UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE QUÍMICA E BIOQUÍMICA

MOLECULAR DIAGNOSIS OF FAMILIAL HYPERCHOLESTEROLEMIA AND FUNCTIONAL CHARACTERIZATION OF MISSENSE VARIANTS IN THE *LDLR* GENE

Sílvia Valentim de Azevedo

Dissertação
Mestrado em Bioquímica
Orientação: Doutora Mafalda Bourbon
2014/2015

Agradecimentos

Primeiramente, quero agradecer à Doutora Mafalda Bourbon por me ter aceite, motivado e ensinado que o trabalho e o esforço compensam sempre. Graças a si e à sua paixão pela investigação sou hoje uma aluna mais capaz e dedicada. É a responsável pelo início do Estudo Português de Hipercolesterolemia Familiar, do qual fazem parte dezenas de médicos e centenas de doentes, aos quais deixo também o meu agradecimento.

À Patrícia, que tanto me ensinou e tantas *brainstorms* teve comigo. Foi, sem dúvida e sem sequer se aperceber, a minha maior inspiração no decorrer deste projecto. Quando for grande quero ser como tu!

À nossa grande família do GIC. Tanto aos que foram passando como aos que ainda ficam, pois tornaram todos os dias aqui únicos e animados. À Salima, Rita, Cibelle e Joana Ch., o meu obrigada. Um agradecimento especial à Ana, pela boa disposição constante, pelo seu interesse e curiosidade naturais; e outro à Catarina, por toda a ajuda e pelas desventuras vividas em Bilbau. Um gigante obrigada à minha Juca, pela partilha de histórias e sentimentos que levaram ao nascer de uma boa amizade, pelas horas de confidências, por tornar todas as horas passadas no laboratório mais felizes.

Ao César por toda a ajuda e orientação em Bilbau. Ao Asier, ao Aitor, à Rocío e à Garazi, por toda a animação e boa disposição que os dias cinzentos teimavam em contrariar. Às minhas colegas de casa Águeda e Angela. Fui super bem recebida e vocês foram, sem dúvida, os responsáveis pelos meses espetaculares que passei nesta cidade fantástica.

Ao meu núcleo de amigos, que nos últimos cinco anos me acompanharam neste percurso. Aos rapazes João, Frederico, Tomás, Gui e Lucas por todo o ânimo; ao Bustorff e ao Coelho pela paciência e por serem bons ouvintes, mesmo quanto 90% das frases que dizia eram constituídas por palavrões de fúria. Às minhas amigas do coração, Inês, Ana, Leonor, Madalena e Margarida, por todos os momentos, em especial pela mágica viagem a Berlim. Estiveram sempre presentes e contribuíram tanto para a minha formação e crescimento enquanto pessoa. Dizem que os amigos da faculdade são para sempre, e espero que sejam mesmo. Obrigada, sem vocês não teria sido o mesmo.

Às pessoas que são a presença mais constante desde há 13 anos, Joana e Andreia. O tempo tem-me trazido muitas coisas e levado outras tantas, mas vocês permanecem. Um obrigada gigante por todas as choradeiras e gargalhadas, por todas as horas a fio passadas à conversa, mesmo quando nada mais havia por dizer. À Joana por me ouvir vezes sem conta e por fazer de mim cúmplice de todos os seus esquemas… bom, nossos esquemas. À Andreia, por estar sempre presente e por todos os conselhos em alturas de, como o seu pai diria, "Oh meu deus, há dramas no mundo encantado das *queens*!". Foi tão bom crescer convosco!

Ao Miguel, quem torna tudo mais fácil, menos dramático e até caricato. Ainda bem que te tenho comigo!

Deixo ainda aos meus pais o maior agradecimento de todos, pois não podiam ter-me amado mais. Sinto que tenho demasiada sorte em ter-vos. Uma vez li que a maior dificuldade que os pais enfrentam é deixar um filho ser ele próprio, e vocês deixaram. Sempre. Muito obrigada. Termino com um agradecimento gigantesco aos meus queridos irmãos Patrícia e Rodrigo, simplesmente por existirem, a vida é tão melhor ao vosso lado!

Sou uma pessoa verdadeiramente feliz.

"Somos o resultado dos livros que lemos, das viagens que fazemos
e das pessoas que amamos." (Airton Ortiz)

Abstract

Familial Hypercholesterolemia (FH) is an autosomal dominant disorder characterized by increased levels of plasmatic cholesterol since birth.

FH occurs due to functional variants in one of three genes: low density lipoprotein receptor (*LDLR*), apolipoprotein B-100 (*APOB*) and proprotein convertase subtilisin/kexin type 9 (*PCSK9*).

This disorder has an estimated prevalence of 1:500 individuals, being more than 90% of FH patients identified with a variant in the *LDLR* gene. The LDLR is a membrane glycoprotein, responsible for binding and uptake of low density lipoproteins (LDL), the major cholesterol transporter in blood.

Variants in *LDLR* gene may result in a defective LDL catabolism, leading to increased cholesterol levels in plasma, accumulating in tendons and arteries. Cholesterol accumulation in untreated FH patients leads to premature atherosclerosis and cardiovascular disease development.

Although more than 1600 variants identified in this gene are reported in databases, the majority of them remains, until now, without functional studies proving their pathogenicity. For patients carrying these variants, a definitive molecular diagnosis for FH is not possible.

The main purpose of Portuguese Study of Familial Hypercholesterolemia (EPHF) is to perform the molecular diagnosis of patients who were clinically diagnosed. Criteria used are adapted from the ones of Simon Broome. Furthermore, the performance of functional studies for variants of unknown pathogenicity is imperative, in order to better clarify the molecular basis of this disorder.

During this project, the molecular diagnosis was performed for 25 index cases, participants of Portuguese FH Study. Among all of them, 11 variants were identified in 12 patients. Among these, 2 were novel, being first described in this project. The remaining 9 have been previously reported, although only 7 have been functionally assessed. The search for large rearrangements was performed by Multiplex Ligation-dependent Probe Amplification (MLPA), but no alteration of this type was identified.

The 10 most common *LDLR* variants described during the Portuguese FH Study, to date without functional assessment, were selected. *In silico* assessment was performed using described tools, in order to predict their pathogenicity. Site-directed mutagenesis was successfully performed in a pcDNA3_LDLR plasmid for all variants, being expressed in CHO–ldlA7 cells, lacking endogenous expression of *LDLR*. LDLR expression, binding and uptake were independently assessed by flow cytometry.

Results suggest that among 10 functionally studied variants, 7 cause an impairment in LDLR function. Variants c.1802A>T p.(Asp601Val), c.1876G>A p.(Glu626Lys), c.631C>G p.(His211Asp), c.661G>T p.(Asp221Tyr), c.618_638del p.(Gly207_Ser213del), c.551G>A

p.(Cys184Tyr) and c.1775G>A p.(Gly592Glu) were classified as pathogenic. However, variants c.1816G>T p.(Ala606Ser), c.1966C>A p.(His656Asn) e c.2177C>T p.(Thr726Ile) seem to be neutral, not revealing any kind of impact on the LDLR function.

Comparing functional studies with *in silico* tools predictions led to the conclusion that these are useful, but should not be the only source of evidence for a diagnosis associated to a pathology.

FH is a disease for which, fortunately, a genetic diagnosis and therapeutic options exist. Although it remains as a subdiagnosed disorder, the performance of a molecular diagnosis, along with functional assessment of variants with unknown pathogenicity, allows a definite diagnosis. This way, preventive measures and personalized counseling can be made in order to improve FH patients' prognostic, providing them longer and better lives.

Key words: Familial hypercholesterolemia, LDLR, molecular diagnosis, variants, functional studies.

Resumo

A Hipercholesterolemia Familiar (FH) é uma patologia genética que é transmitida de forma autossómica dominante e é caracterizada por elevados níveis de colesterol no plasma desde o nascimento.

A FH ocorre devido a variantes funcionais num dos genes codificantes de três proteínas: recetor de lipoproteínas de baixa densidade (*LDLR*), apolipoproteína B-100 (*APOB*) ou pró-proteína convertase subtilisina quexina tipo 9 (*PCSK9*).

O LDLR é uma glicoproteína membranar que liga e internaliza colesterol associado às lipoproteínas de baixa densidade (LDL), que constituem o principal transportador de colesterol no sangue. Variantes no gene *LDLR* podem resultar num catabolismo deficiente das LDL, tendo como consequência o aumento do colesterol no plasma, que se acumula nos tendões e artérias. Esta acumulação pode levar ao desenvolvimento prematuro de aterosclerose e doença cardiovascular.

A FH apresenta duas formas clínicas: a forma heterozigótica, que apresenta um fenótipo menos agressivo, com valores de colesterol total entre 290 e 500 mg/dl (com LDL>190 mg/dl); e a forma homozigótica, que apresenta um fenótipo mais agressivo, com valores de colesterol total entre 600 mg/dl e 1000 mg/dl.

Estima-se que as prevalências das formas heterozigótica e homozigótica sejam de 1/500 e de 1/1000000 indivíduos, respetivamente. Sabe-se ainda que mais de 90% dos casos identificados apresentam uma variante no *LDLR*, fazendo deste gene o mais associado a esta doença.

Apesar de haver mais de 1600 variantes do *LDLR* reportadas em bases de dados, para a maior parte delas não existe registo de estudos funcionais que provem a sua patogenicidade, levando a que não seja possível atribuir um diagnóstico molecular definitivo a estes casos. Assim sendo, a necessidade de avaliar funcionalmente estas variantes, de modo a perceber em que medida afetam a função do recetor, também se torna imperativa.

Tendo em conta as prevalências acima referidas, estima-se que em Portugal existam cerca de 20000 casos de FH. O Estudo Português de Hipercolesterolemia Familiar (EPHF), implementado em 1999 no Instituto Nacional de Saúde Dr. Ricardo Jorge, tem como objectivo a determinação da prevalência e distribuição desta patologia em Portugal. A população em estudo é constituída por indivíduos de ambos os sexos e todas as idades, desde que cumpram com os critérios para o diagnóstico clínico de FH.

O diagnóstico clínico de FH é feito, em Portugal, de acordo com os critérios clínicos adaptados de *"Simon Broome Heart Research Trust"* e idealmente deverá ser realizado o estudo genético, com a identificação da variante, pois só desta forma é possível confirmar o diagnóstico clínico. A realização de estudos funcionais para determinar o efeito de

variantes identificadas na função da proteína é também de elevada importância, podendo contribuir para uma terapêutica mais personalizada.

O EPHF está dividido em duas partes: o estudo bioquímico e o estudo molecular. Este último está sub-dividido em cinco fases. A fase I compreende a extração de DNA, o estudo do promotor, todos os exões e regiões adjacentes do LDLR, bem como das variantes mais comuns nos exões 26 e 29 da *APOB*. A fase II consiste no estudo de grandes rearranjos por *Multiplex Ligation-dependent Probe Amplification* (MLPA). A fase III consiste no estudo do *PCSK9* e a fase IV no estudo de todo o gene da *APOB*. A V e última fase consiste na realização de estudos funcionais *in vitro* para variantes cuja patogenicidade ainda é desconhecida.

Assim sendo, este projeto está dividido em duas partes, que compreendem três das cinco fases do EPHF, nomeadamente as fases I, II e V.

A primeira parte consistiu na realização do estudo molecular em participantes do EPHF, estudando o promotor e os 18 exões e regiões adjacentes do *LDLR,* bem como o estudo de parte dos exões 26 e 29 da *APOB* por *Polymerase Chain Reaction* (PCR) e sequenciação de Sanger. De seguida, procedeu-se à pesquisa de grandes rearranjos por MLPA. Uma predição, utilizando as ferramentas *in silico,* foi também realizada, de modo a prever o impacto das alterações encontradas ao nível da proteína. Esta predição foi realizada tanto para alterações identificadas ao nível do exão, com as ferramentas *Polymorphism Phenotyping (PolyPhen-2), Sorting Tolerant From Intolerant* (*SIFT*) e *Mutationtaster*, como para as alterações identificadas ao nível do intrão, de modo a prever efeitos no *splicing*, com as ferramentas *Human Splicing Finder (HSF), the Splice Site Prediction by Neutral Network (NNSSP)* e *FSPLICE*.

A segunda parte deste projeto consistiu então no estudo funcional das 10 variantes do *LDLR* mais comuns na população portuguesa, até à data sem estudos funcionais. As diferentes variantes do *LDLR* foram obtidas por mutagénese dirigida num plasmídeo pcDNA3_LDLR sob o controlo do promotor viral SV40. Toda a região de interesse foi confirmada por sequenciação de Sanger e foi feita uma reclonagem, em que o gene *LDLR*, já com a variante, foi transferido para um vetor limpo. Células CHO-ldlA7, que não expressam endogenamente o recetor, foram transfetadas com os diferentes plasmídeos. A expressão do recetor foi avaliada através da deteção com anticorpos; a ligação e a internalização foram avaliados através do uso de LDL fluorescentemente marcada com FITC. O impacto de todas as variantes ao nível da expressão, ligação e internalização foi analisado por citometria de fluxo.

Em 25 casos índex estudados, 11 variantes foram identificadas em 12 doentes, embora duas destas sejam provavelmente benignas. A variante patogénica mais frequentemente encontrada na *APOB* foi identificada em apenas 2 doentes. De entre as alterações encontradas no LDLR (10), 2 foram aqui primeiramente reportadas e 8 já tinham sido anteriormente identificadas. De entre estas últimas, 7 já apresentavam estudos funcionais comprovando a sua patogenicidade.

Aquando da pesquisa por grandes rearranjos nestes doentes por MLPA, nenhuma alteração deste tipo foi identificada no grupo em estudo.

O estudo funcional das dez alterações mais frequentemente identificadas no decorrer do EPHF, até à data sem estudos funcionais que comprovem o impacto na função do recetor, revelou que, de entre as 10 variantes estudadas, 7 são patogénicas, afetando de alguma forma a função do LDLR.

A variante c.1802A>T p.(Asp601Val) é patogénica, fazedo com que não haja sequer LDLR à superfície celular. Como não atinge a membrana (por não ser expressa ou por não se ancorar a esta) apresenta valores de ligação e de internalização igualmente baixos. As variantes c.1876G>A p.(Glu626Lys), c.631C>G p.(His211Asp), c.661G>T p.(Asp221Tyr), c.618_638del p.(Gly207_Ser213del) e c.551G>A p.(Cys184Tyr) apresentaram uma expressão normal. No entanto, foram observáveis valores muito reduzidos para a fluorescência associada à união das LDL nestes casos. Consequentemente, a internalização também é defetiva, parecendo lógico classificar estas variantes como patogénicas. A variante c.1775G>A p.(Gly592Glu) resulta num defeito, possivelmente, ao nível da reciclagem do recetor, sendo também considerada patogénica. Por outro lado, as variantes c.1816G>T p.(Ala606Ser), c.1966C>A p.(His656Asn) e c.2177C>T p.(Thr726Ile) parecem não revelar qualquer impacto no LDLR, sugerindo a sua neutralidade, pois obtiveram-se valores de fluorescência, associados à expressão e atividades de ligação e internalização, comparáveis ao wt. Estes resultados sugerem que os doentes portadores destas variantes devem apresentar outra justificação para o seu fenótipo hipercolesterolémico.

A análise destas variantes através de ferramentas de predição *in silico*, realizada para todas as variantes identificadas no decorrer deste projeto, permitiu concluir que estas predições nem sempre vão de encontro ao determinado através da realização de estudos funcionais. Assim sendo, estes programas deverão ser usados, mas com a devida ressalva de que são apenas preditores.

A FH é uma doença para a qual, felizmente, existe um diagnóstico definitivo (genético) e variados tratamentos farmacológicos. Apesar de ser uma doença subdiagnosticada, estão a ser realizadas diversas iniciativas para que haja uma maior divulgação da doença, nomeadamente dos benefícios do diagnóstico precoce. No âmbito da EPHF tem-se feito um esforço para que seja implementado o diagnóstico molecular como diagnóstico preferencial, juntamente com a execução de estudos funcionais para determinar a patogenicidade de variantes desconhecidas. Só assim os doentes têm a possibilidade de ter um diagnóstico definitivo para a sua patologia tornado possível instituir medidas preventivas com uma terapêutica dirigida e personalizada, melhorando o prognóstico destes doentes.

Palavras chave: Hipercolesterolemia familiar, LDLR, diagnóstico molecular, variantes, estudos funcionais.

Abbreviations

Aa	Amino acid
ACAT	Acyl-CoA:Cholesterol acyltransferase)
ARH	Autosomal Recessive Hypercholesterolemia
Apo	Apolipoprotein
bp	Base pair
cDNA	Complementar DNA
CHD	Coronary Heart Disease
CHO	Chinese Hamster Ovary
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	Desoxirribonucleótido trifosfatado
EDTA	Ethylenediamine tetraacetic acid
EGF	Epidermal Growth Factor
EPHF	Estudo Português de Hipercolesterolemia Familiar
ER	Endoplasmic Reticulum
FACS	Fluorescence Activated Cell Sorter
FH	Familial Hypercholesterolemia
FITC	Fluorescein isothiocyanate
g	gram
g	Relative centrifugal force
HDL	High Density Lipoprotein
HDL-C	HDL Cholesterol
HGVS	Human Genetic Variation Society
HMGCoA	3-hydroxy-3-methylglutaryl coenzyme A
IC	Index case
IDL	Intermediate Density Lipoprotein
INSA	Instituto Nacional de Saúde Dr. Ricardo Jorge
kb	Kilobase
kDa	Kilodalton
LDL	Low Density Lipoprotein
LDL-C	LDL Colesterol
LDLR	Low Density Lipoprotein Receptor
LDLRAP 1	LDLR adaptor protein 1
LPL	Lipoproteín lipase
Lp(a)	Lipoproteína (a)
M	Molar concentration

mg	Miligram (10^{-6} g)
MI	Miocardial Infarction
MLPA	Multiplex ligation-dependent probe amplification
mM	Milimolar (10^{-3} M)
mRNA	Messenger Ribonucleic acid
ng	Nanogram (10^{-9} g)
NGS	Next Generation Sequencing
NNSSP	Nearest-neighbor Secondary Structure Prediction
PBS	Phosphate buffered saline
PCR	Polymerase Chain Reaction
PCSK9	Proprotein convertase subtilisin/kexin type 9
PTCA	Percutaneous transluminal coronary angioplasty
RNA	Ribonucleic acid
RT	Room temperature
SAP	Shrimp Alkaline Phosphatase
SB	Simon Broome
SDS	Sodium dodecyl sulfate
SIFT	Sort Intolerant From Intolerant
TBE	Tris-borate-EDTA
TG	Triglyceride
U	Enzyme unit
VLDL	Very Low Density Lipoprotein
VUS	Variant of Uncertain Significance
V/V	Volume/volume
WT	Wild type
W/V	Weight/volume
ºC	Celsius degree
µg	Microgram (10^{-6} g)
µL	Microliter (10^{-6} L)
ρmol	Picomol (10^{-12} mol)

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1. Cholesterol and lipoproteins

Cholesterol is a very important molecule in biology, which led to the awarding of several Nobel prizes to scientists who devoted part of their careers to its study [1][2][3][4]. It is an organic molecule of extreme importance in mammalian cells, being an integral component of cell membranes, due to its insolubility. Furthermore, it is essential for steroid hormone synthesis, bile acid metabolism, and as a building block for cellular platforms such as lipid rafts [5]. However, its insolubility has as much advantages as disadvantages – cholesterol is capable of creating a barrier to protect the cell through regulation of its interaction with the exterior but, when it accumulates within the wall of an artery, it cannot be readily mobilized, and its presence eventually leads to the formation of an atherosclerotic plaque [5].

Cholesterol is biosynthesized in all animal cells through the mevalonate pathway, being the mevalonate production the rate-limiting and irreversible step of its biosynthesis [5][6]. This important step is performed by the enzyme Hydroxymethylglutaryl-CoA (HMG CoA) reductase, which constitutes a target for cholesterol-lowering drugs [7]. In addition to biosynthesis, cholesterol is obtained through the diet, being withdrawn from the intestinal lumen, through the intestinal epithelial cells, and reaching the blood flow.

In mammals, the cholesterol transport is facilitated by esterifying the sterol with long-chain fatty acids and packaging these esters within the hydrophobic cores, inside plasma lipoproteins [5].

Lipoproteins can be separated by ultracentrifugation based on their densities (*table 1.1.*) and have been categorized into six major classes: chylomicrons, Very-Low-Density Lipoprotein (VLDL), Intermediate-Density Lipoprotein (IDL), Low-Density Lipoprotein (LDL)

and High-Density Lipoprotein (HDL) and Lipoprotein(a) (Lp(a)) [8]. All proteins play major roles in mammals' organism [9][10], as they are the biological mediators of cholesterol and triglycerides transport, taking part of numerous processes in several pathways. However, LDL is the most prominent lipoprotein in plasma as well as the primary plasma carrier of cholesterol, being responsible for its delivery to all tissues [11].

Table 1.1. Characteristics and percentage content of the distinct lipoprotein particles. *Adapted from [12].*

Lipoprotein particle	Size (Å)	Density	C %	TG %	PL %	ApoP %	Major apoproteins
Chylomicrons	800–5000	0.95	3	90	5	9	AI, AII, B, CI, CII, CIII
VLDL	300–800	0.95–1.006	10	70	10	10	BI, CI, CII, CIII, E
IDL	250–350	1.006–1.019	–	–	–	–	B, CIII, E
LDL	180–280	1.019–1.063	26	10	15	25	B
HDL	50–120	1.063–1.210	20	5	25	50	AI, AII

C, cholesterol; TG, triglycerides; PL, phospholipids; ApoP, apoprotein.

Each LDL particle has a diameter of about 22 nm and a mass of 3000 kDa, containing approximately 1500 molecules of cholesteryl ester in a hydrophobic core surrounded by a polar phospholipid coat and a single large protein called apolipoprotein B (apoB) [13] (*Figure 1.1.*).

LDL is not secreted directly from the liver, but rather produced in the circulation from VLDL, which is secreted by the liver and transports mainly triglycerides to adipose tissue and muscle. After the removal of these triglycerides in capillaries, it is transformed in IDL. Some of IDL particles, which have apolipoprotein E (apoE) and apoB-100 in their constitution, are rapidly taken up by the liver; others remain in the circulation, where they undergo further triglyceride hydrolysis, lose their apoE particles and are converted to LDL [14].

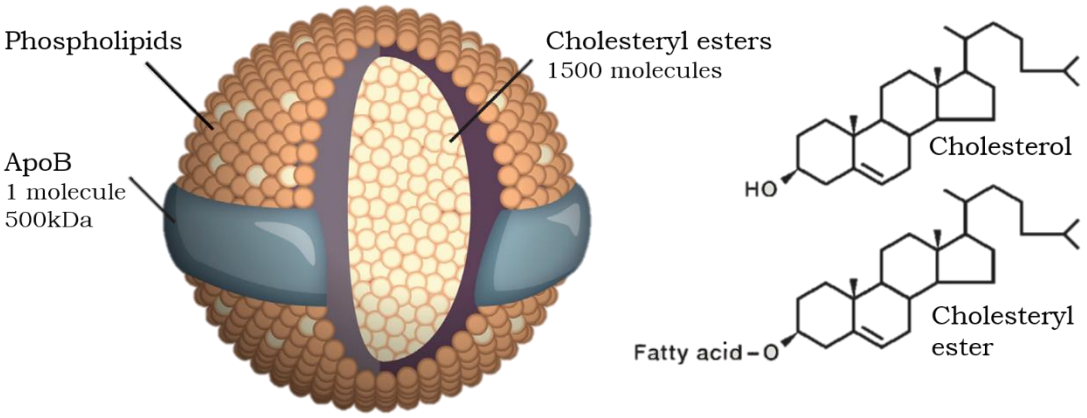


Figure 1.1. LDL, a cholesterol transporter*. LDL is a spherical particle with 220 nm of diameter and a mass of 3000 kDa. Each particle contains approximately 1500 molecules of cholesteryl ester in the hydrophobic core and a hydrophilic coat composed of phospholipids, unesterified cholesterol molecules and 1 molecule of apoB. Adapted from [7].*

The packaging of cholesterol into lipoproteins allows its correct transport to cells. However, as cholesteryl esters are too hydrophobic to pass through membranes, this delivery problem is solved by lipoprotein receptors, namely the LDL receptor (LDLR), through which approximately two thirds of LDL clearance is normally mediated [5].

1.1.1. The LDLR pathway

The LDLR is a cell membrane glycoprotein which is ubiquitously expressed. Nevertheless, the largest number of LDLR is produced by the liver [5]. The LDLR binds two proteins: apoB-100, the 387 kDa glycoprotein that is the sole protein of LDL, and apoE, a 34kDa protein that is also found in multiple copies in IDL [5].

The LDLR is synthesized in the rough endoplasmic reticulum (ER) as a precursor with an apparent molecular weight of 120 kDa. Posteriorly, it migrates to the Golgi apparatus, where it undergoes extensive glycosylation, reaching the mature form of 160 kDa [15].

After synthesis, LDL receptors appear on the cell surface, where they gather in coated pits, ready to perform the receptor-mediated endocytosis [5] *(figure 1.2.)*. The receptor binds a LDL particle, due to its affinity to apoB, and the coated pits invaginate to form coated endocytic vesicles [5]. Here, depending on the cell type, specific proteins (e.g. LDLR adaptor protein 1 (LDLRAP1) and Disabled homolog 2 (Dab2) [16]) (not shown in *figure 1.2.*) play an important role because they bind the cytoplasmic tail of LDLR and governs its clustering into clathrin-coated pits, being required for internalization of the LDL-LDLR complex and

for efficient binding [17]. Very quickly, the clathrin coat dissociates and multiple endocytic vesicles then fuse to create endosomes, where the LDLR separates from the LDL due to the acid pH, created by ATP-driven proton pumps [18]. A segment of the endosomal membrane forms a recycling vesicle, responsible of returning the LDLR to the cell surface. The endosome containing LDL fuses with a lysosome and its protein content is hydrolysed to amino acids and cholesteryl esters are hydrolysed to cholesterol. The liberated cholesterol is used to cellular functions as plasma membranes, bile acids and steroid hormones synthesis; or stored in the form of cytoplasmic cholesteryl ester droplets until further use by the cell. The receptor can be recycled several times, since one round trip lasts 10 minutes and LDLR has a 20-hour lifespan [5].

The LDLR also has affinity to proprotein convertase subtilisin/kexin type 9 (PCSK9) [19], secreted from hepatocytes. Similarly to the LDL, PCSK9 can also interact with LDLR, forming a tight ligation impairing, in the endosome, the recycling of the receptor to cell surface and targeting LDLR to the lysosome for degradation [20]. Thus, PCSK9 is also a modulator of the LDLR pathway.

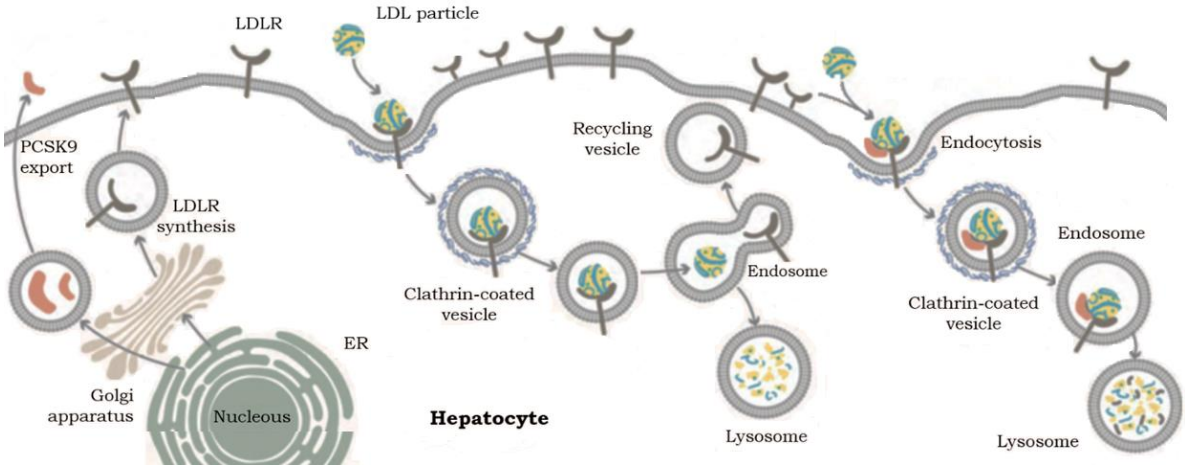


Figure 1.2. The LDLR pathway *- The LDLR is synthesized in the ER, undergoes extensive glycosylation in the Golgi apparatus and it transported to the cell surface. The LDLR specifically binds ApoB in LDL particles, internalizing them by endocytosis. Inside the endosome the complex dissociates; the receptor is recycled to the cell surface, whereas the LDL particle is degraded into the lysosome. The PCSK9 is synthesized and excreted, binding LDLR–LDL complex extracellularly under regulatory signals. PCSK9 prevents the dissociation of LDLR-LDL complex, leading to its degradation in the endosomal compartment. Adapted from [21].*

The LDLR is a crucially important modulator of plasma LDL levels as it removes not only LDL but its precursors IDL from circulation due to the presence of apoE and apoB, whereby the receptor binds [1][20]. Its high affinity for LDL and ability to cycle multiple

times in and out of the cell allow the delivery of large amounts of cholesterol to body tissues. In addition, both cholesterol biosynthesis and LDLR pathway are regulated by negative feedback in order to keep the level of cholesterol in cell membranes constant, through regulation of HMG CoA reductase, and LDLR transcription factors [1],[20].

When a defect arises in genes codifying one of the proteins necessary to the performance of the LDLR pathway, this can lead to a lipid disorder known as Familial Hypercholesterolemia (FH), or Autosomal Recessive Hypercholesterolemia (ARH) in case of defects in LDLRAP1. It was the study of FH by several scientists which threw light on fundamental biological and regulatory mechanisms related with cholesterol.

1.2. Familial Hypercholesterolemia

The correlation between plasma cholesterol and coronary heart disease (CHD) was first postulated through the description of families in which high plasma cholesterol levels and its associated coronary problems were transmitted as an autosomal dominant trait [22]. Later, it was found that these families had a genetic disorder of lipid metabolism - Familial hypercholesterolemia (FH).

FH is a common autosomal dominant characterized by high levels of low density lipoprotein cholesterol (LDL-C) in plasma and increased risk of premature coronary heart disease (CHD) [5]. FH has an estimated prevalence of 1/500 individuals, although a prevalence of 1/200 has been observed in some populations [23], which leads to a calculated value between 14 and 34 million affected individuals worldwide, being among the commonest inherited disorders [24].

Clinically, FH exists essentially in two forms: a more common and less severe heterozygous form, and a rare (prevalence 1/1000000) and more severe homozygous form [5][15]. FH heterozygotes have a two-fold increase in the number of LDL particles in plasma from the time of birth, which predisposes to premature CHD as early as 30 years old [5]. Homozygotes patients have a much worse prognosis, presenting six to ten-fold elevations in plasma LDL levels from the time of birth, often having a myocardial infarction in childhood [5].

1.2.1. Therapeutic approaches for FH

Although cholesterol is endogenously synthesized, the diet is also a source of cholesterol, and these are two points where therapeutic measures can be applied. FH patients should follow a strict diet, poor in fat and perform daily physical activity, but these measures are not enough to lower their increased LDL-C values. Fortunately, FH is a disorder for which effective treatment exists and current therapies revolve around

cholesterol-lowering drugs and in some cases LDL apheresis [25]. For FH patients, medication should be initiated as soon as possible still in childhood, being strongly considered starting between 8 and 10 years old [23][26].

Several lipid lowering drugs are currently available due to knowledge of fundamental properties of the cholesterol metabolism, as well as LDLR and its interactions with other important molecules as APOB and PCSK9. These interactions represent central implications for therapy of FH.

The first inhibitor of HMG CoA reductase, and consequently inhibitor of the endogenous cholesterol synthesis, initiated the class of cholesterol-lowering drugs known as statins [27]. This drug leads to the synthesis of more LDLR, in order to satisfy the cell demands for cholesterol. The final outcome is the reduction of LDL-C between 20-45%, depending on the dosage [28]. Nonetheless, statins do not lower LDL significantly in FH homozygotes (HoFH), who have null variants in both copies of the LDLR [29].

Statins are often coadministered with ezetimibe [23], a cholesterol absorption inhibitor that actuates at the level of small intestine, reducing the amounts of dietary cholesterol that reach the liver. This results in increased LDL withdrawn from the circulation. Ezetimibe might as well be administrated alone in adults, in case of intolerance to statins, resulting in a reduction of approximately 20% of the LDL-C levels [28][29][30][31].

For a long time, statins and ezetimibe were the only pharmacological treatments available for FH [32]. However, management of homozygous FH requires additional treatment as LDL apheresis, which provides transient reductions in LDL-C levels by 40% [33][34]. LDL apheresis is typically performed once or twice-a-week in patients with homozygous FH and is also an option for the treatment of heterozygous FH, intolerant to statins. Nonetheless, this invasive therapy is far from being the ideal, since it has several side effects associated and the treatment sessions are expensive and time consuming [21].

Advances in genetic-based pharmacology have empowered the study of new LDL-lowering agents, which are currently at advanced stages of development. These comprise the development of monoclonal antibodies targeting PCSK9, an anti-sense oligonucleotides targeting APOB and cholesteryl ester transfer protein inhibitors [23]. Studies are underway to determine the long-term safety of these therapeutic measures and their efficacy in preventing CHD [35][36]. Nevertheless, further studies on these new medications' long-term safety and efficacy are still needed.

It is now clear that LDLR pathway is of extreme importance in regulation of cholesterol levels in blood and its activity has an impact on the response of the system to lowering the levels of cholesterol. However, although FH has a high prevalence and a clear relation with premature atherosclerosis and CHD, it remains as an extremely underdiagnosed disorder worldwide [23], which emphasises the need of investigation at the genetic level. Only the molecular diagnosis can confirm a clinical diagnosis of FH, hence

the genetic demonstration of a causative variant in specific genes is important for FH diagnosis, allowing a more personalised treatment [23].

1.2.2. Genetics behind FH

Genetically, heterozygous FH is caused by variants in three genes: *LDLR*, *APOB*, and *PCSK9* [5][37][38].

Variants in the *LDLR* gene are the most common, being the cause of more than 90% of the identified FH cases worldwide [39]. These variants can result in an impaired function of the LDLR, which will be further discussed in section 1.2. Molecular FH.

Variants in *APOB* sequence, affecting amino acids which are important for the binding to LDLR, can disable the recognition of LDL by the LDLR [40]. This results in a reduction of LDL withdrawn from the circulation, leading to increased cholesterol levels. Due to *APOB* size, only part of exons 26 and 29 is routinely studied, although disease-causing variants have been reported out of these sites [41].

PCSK9 gain-of-function variants result in disruption of the recycling mechanism, which is responsible for the LDLR return to cell surface after internalization, as the PCSK9 targets the complex LDLR-LDL to lysosomal degradation. As a result, the number of cell-surface LDLR declines and LDL rises [42][43]. *PCSK9* loss-of-function variants have as well been reported, resulting in an enhancement of the recycling process, consequently reducing LDL-C levels [44][45].

Variants in *APOB* and *PCSK9* genes are found in ~5% and ~1%, respectively, of heterozygous FH subjects with a causative variant [40]. Homozygous individuals are uncommon and present a more aggressive clinic phenotype, leading to extremely increased cholesterol values, since there is no production of normal and functional protein. Some rare subjects are “double heterozygotes”, which means they carry variants in two of the above-mentioned genes.

Recently, variants in *LDLRAP1* were identified as the cause of autosomal recessive hypercholesterolemia (ARH) [46], a disorder distinct from FH as it causes a less severe phenotype and these patients are more responsive to lipid lowering therapies. The protein LDLRAP1 facilitates the internalization in clathrin coated-pits, thus variants causing its loss of function can result in reduced clearance of circulating LDL by the liver [47], leading to a phenotype between heterozygous and homozygous FH [48].

A genetic defect in one of the three FH-related genes will lead to defects in distinct steps of the LDLR pathway, depending on the affected gene, which can culminate in premature cardiovascular disease, as represented in *figure 1.3*.

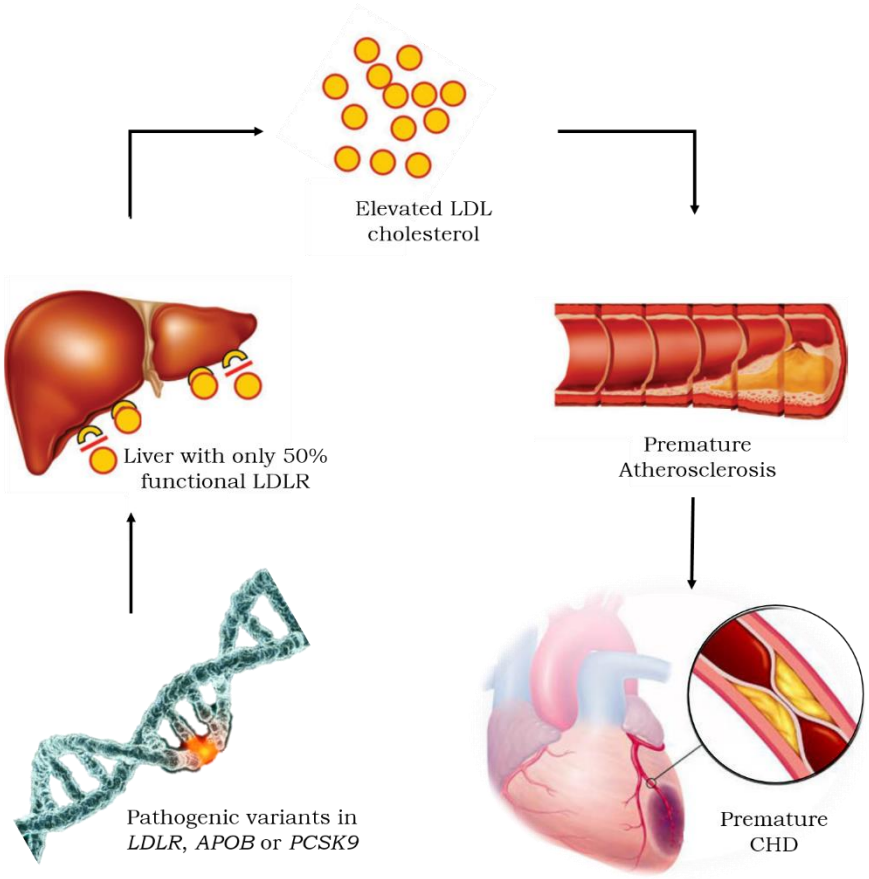


Figure 1.3. - Pathophysiology of Familial Hypercholesterolemia*. A heterozygous disease-causing variant in one of the three genes associated with FH leads to a situation where the liver only produces 50% of functional LDLR, resulting in elevated LDL-C. Subsequently premature atherosclerosis can urge, culminating in CHD. Adapted from [23].*

In view of genetic variability and when lifelong drug treatment is under consideration, the comprehension of the origin of a *LDLR* defect is imperative. *LDLR* variants represent the most known cause of FH, thus knowledge on the *LDLR* gene and protein will be covered in next section, as it allows profound understanding of the molecular basis of this disorder.

1.3. Molecular FH

1.3.1. The *LDLR* gene and protein

The striking feature of the LDLR pathway is that it requires a highly selective and distinct movement for each one of the components involved, characteristics that may reside in its structure. These domains are encoded in the *LDLR* gene, which lies on the short arm

of chromosome 19 - 19p13 -, spans 45kb, and is comprised of 18 exons and 17 introns [15]. The protein coding sequence is interrupted by introns in such a way that many of the protein segments are revealed as products of individual exons [49] *(Figure 1.4.)*.

The LDLR is synthesized as a precursor of 860 amino acids. Nevertheless, the first 21 amino acids, at the extreme NH_2 terminus, constitute a typical hydrophobic signal sequence that is cleaved from the protein prior to its appearance on the cell surface. The short 5' untranslated region plus the signal sequence of the protein is encoded by exon 1 [49]. The mature form of LDLR has 839 amino acids with five recognizable domains [49][50], described below.

The first domain of the LDLR consists of the NH_2-terminal 292 amino acids and is assembled from multiple repeats of 40 residues each. Each repeat has six cysteine residues, all involved in disulphide bonds [51], which must be the cause of the extreme stability of the binding domain of the receptor. All of the charged residues that are conserved bear a negative charge, which might be responsible for the LDLR ability to bind closely spaced positively charged residues. It contains the binding site for apoB and apoE. The repeats I, III, VI and VI are encoded exons 2, 3, 5 and 6, respectively. The other three repeats (III, IV and V) are all contained in a single exon, the exon 4 [49].

The second LDLR domain, constituted by approximately 400 amino acids, is homologous to a portion of the extracellular domain of the Epidermal Growth Factor (EGF) precursor. This part of the LDLR is implicated in the release of bound lipoproteins at low pH in the endosome [18]. This EGF precursor homology domain contains three repetitive sequences of about 40 amino acids, that are designated A, B and C, containing each one six cysteine residues spaced at similar intervals. Each repeat is contained within a single exon – exons 7, 8 and 14. The exons 9 to 13 codify for the region of β-propeller, between the repeats B and C [49].

The third domain is encoded within the single exon 15 [49], consisting of a stretch of 58 amino acids that contains 18 serine or threonine residues, many of which appear to serve as attachment sites for O-linked carbohydrate chains added to serine and threonine residues during posttranslational processing events.

The fourth domain, with 22 hydrophobic amino acids, is the membrane spanning region, which is poorly conserved among species. This transmembrane domain is encoded by exons 16 and 17 [49].

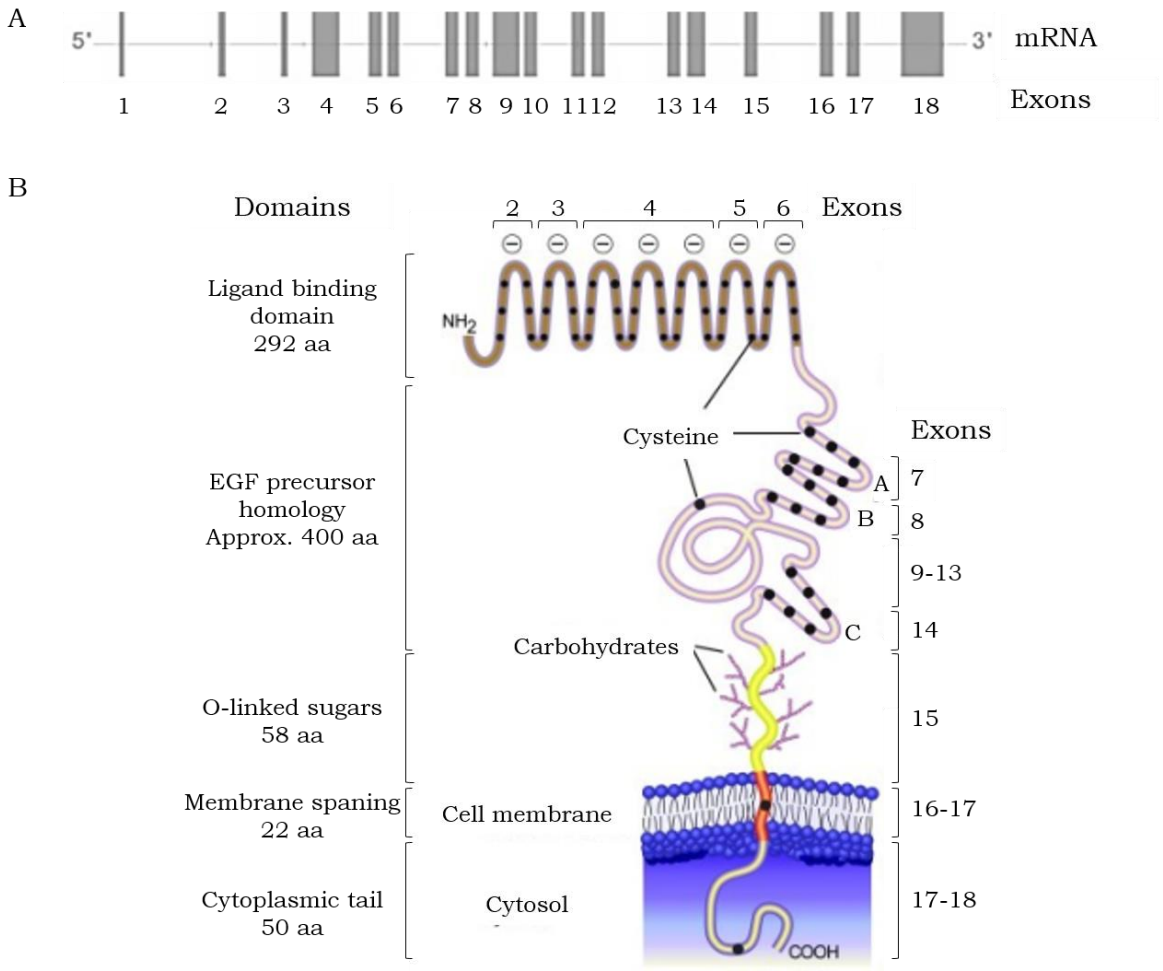


Figure 1.4. Schematic representation of the human LDLR gene (A) and protein (B). *The mature human LDLR is composed of five domains. The 5' untranslated region, as well as the signal peptide are not represented in the scheme. Aa, amino acids. Adapted from [52].*

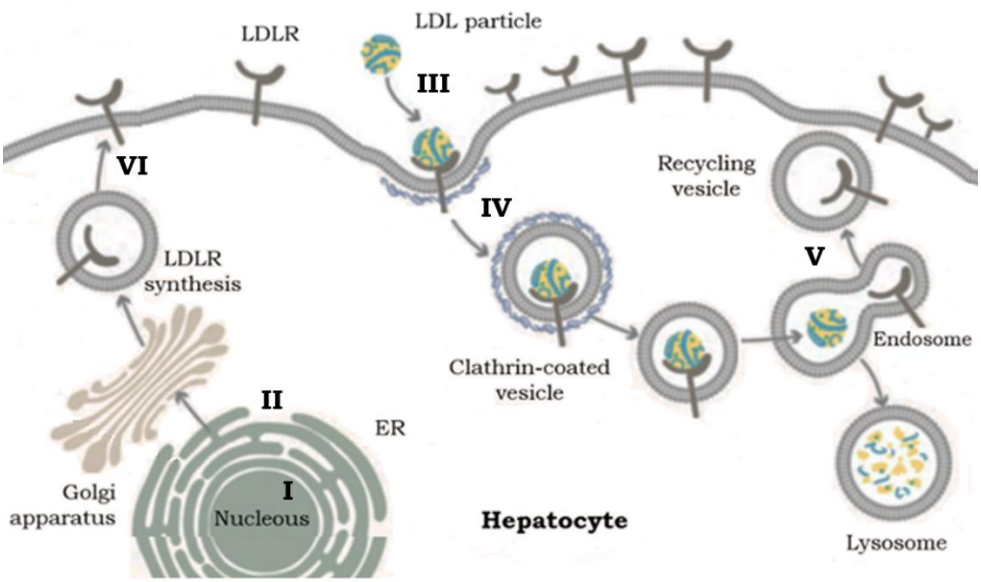
The last protein domain is a 50 amino acid COOH-terminal cytoplasmatic tail, projected into the cytoplasm [49][50]. This domain is strongly conserved among species and plays a crucial role in clustering in coated pits. The cytoplasmic domain is encoded by exons 17 (13 amino acids of the transmembrane domain and the first 39 amino acids of the cytoplasmic domain) and 18, the largest exon in the gene. It encodes the last 11 amino acids of the LDLR and a 2.5 kb DNA sequence that represents the 3' untranslated region of the mRNA [49].

When a change in the LDLR codifying sequence is noted, it might have serious repercussions at the protein and consequently at the LDLR pathway level, implying modifications in the cholesterol level.

According to the domain altered by these modifications at the DNA sequence level, different functions of the LDLR can be affected, which leads us to the LDLR classes of variants.

1.3.2. Classes of variants in LDLR

Consistent with the first reports, there were four classes of LDLR variants [53]. However, posterior new findings brought the necessity to subdivide some of these classes and to create a new one [1][15], existing currently five classes of LDLR variants *(figure 1.5.)*.



Class	Synthesis	Transport ER to Golgi	Binding	Clustering in coated pits	Recycling	Insertion in membrane
I	X					
II	→	X				
III	→	→	X			
IV	→	→	→	X		
V	→	→	→	→	X	
VI	→	→	→	→	→	X

Figure 1.5. Classes of variants that disrupt the structure and function of LDLR*. Each variant affects a different region in the gene resulting in defects in distinct parts of the cycle. Adapted from [5][21].*

Class I Variants: Null Alleles. The variants responsible for this class lead to the production of no LDLR or only trace amounts of it. The absence of receptor protein in these cells may be due to a rapid turnover of the mRNA or to accelerated degradation of the receptor protein [15].

Class II Variants: Transport-defective Alleles. These alleles encode LDLR which is either completely (Class II A) or partially (Class II B) blocked in transport between the endoplasmic reticulum and the Golgi apparatus, not reaching the cell surface. Class II variants comprehend the most common at the *LDLR* locus [15].

Class III Variants: Binding-defective Alleles. These LDLR are normally synthesized, matured and reach the cell membrane. However, their affinity to apoB in LDL or apoE in IDL can be affected, resulting in an impaired binding to LDL [50].

Class IV Variants: Internalization-defective Alleles. Here the receptors move to the cell surface and bind LDL normally. The problem resides in the clustering in clathrin-coated pits, preventing the LDL internalization. These variants have been classified into two groups: variants that alter the cytoplasmic domain alone (Class IV A) and variants that involve the cytoplasmic domain together with the adjacent membrane-spanning region (Class IV B) [15]. This last one produce truncated receptors that lack the membrane-spanning domain as well as the cytoplasmic tail. Most of these molecules are secreted from the cell, but approximately 10% remain adherent to the cell membrane where they bind LDL but do not internalize it.

Class V variants: Recycling-deficient Alleles. The LDLR encoded by these variants perform all steps of the LDLR cycle until the recycling of the receptor, as they fail to release the ligands in the endosome. This results in the degradation of the complex LDLR-LDL, not being recycled to the cell surface [15].

Recently, a sixth class (class VI) has been defended as affecting LDLR insertion in cell membrane [54]. Here should be included variants in the cytoplasmic domain, which makes the anchoring of the receptor to the cell membrane impossible.

1.3.3. Variants

A mutation has been primarily defined as a heritable change in a specific DNA sequence when compared with the reference sequence [55]. However, the mutation within this definition can have an effect at the phenotype level or be a neutral variation without an observable effect. If it is a rare variation, it is called mutation. However, if it occurs in the population at a frequency above 1%, it might be defined as polymorphism [55][56].

Therefore, in order to end these ambiguities, every change in DNA is called variant [56]. Thus, if a dominant phenotype, as in FH, segregates with the variant and does not segregate in its absence, then the variant is compatible with a genotype that can be called disease-causing [55]. These are variants that are likely to have greater functional importance, affecting polypeptide structure and function [55].

The impact of a variant at the protein function level can be assessed with *in silico* tools, allowing a bioinformatics analysis that can be helpful in characterizing new variants found, on which pathogenicity is unknown. However, only the *in vitro* study of the effect of a variant in a protein (functional assay) can correctly determine its pathogenicity.

1.4. Portuguese FH Study

Universal screening for FH was recommended from the World Health Organization (WHO) in 1998, which led to the implementation, at National Institute of Health (INSA) Doutor Ricardo Jorge, of the Portuguese FH Study [57]. The Portuguese FH Study has the purpose of determining the prevalence and distribution of FH in Portugal. The population under study consisted of individuals of both sexes and all ages with a clinical diagnosis of FH (Simon Broome criteria). The implementation of the molecular study of this disorder in Portugal, promotes its early identification in patients and respective relatives, leading to a correct counselling as soon as possible, decreasing their CV risk.

Since 1999, a total of 2122 individuals were enrolled due to the participation of numerous clinicians from several clinics and hospitals in all country. Among these, 623 heterozygous patients had putative pathogenic variants in *LDLR*, 33 in *APOB*, and 4 in *PCSK9* [58].

This study implies a biochemical and a molecular assessment, being the last subdivided into 5 phases, as schematically represented in *figure 1.6*.

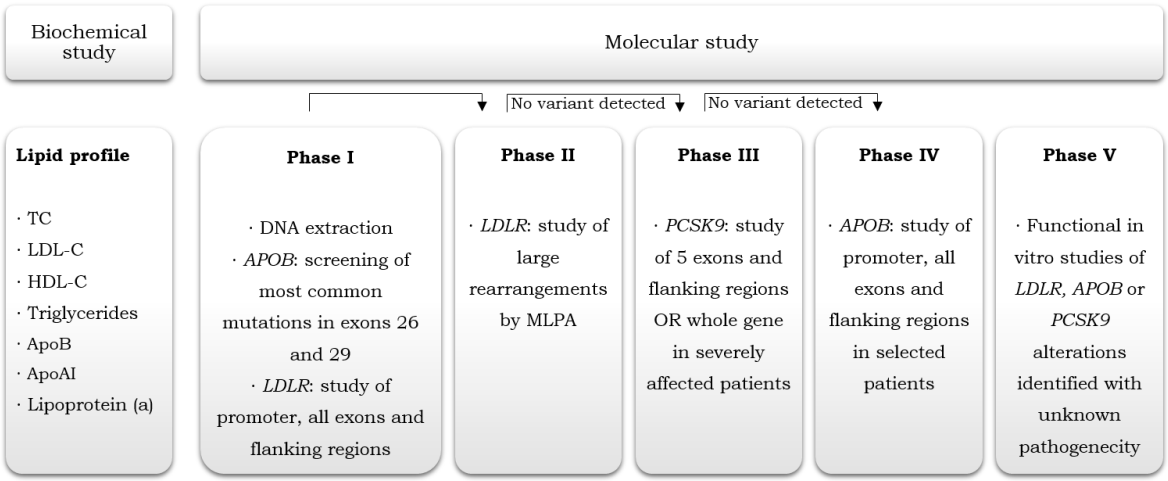


***Figure 1.6. Portuguese FH Study phases**, comprehending the biochemical and the several phases of the molecular study. Adapted from [58].*

When a putative disease-causing variant is identified in an index case, a cascade screening is performed, when samples of relatives are available, as recommended by international guidelines [59]. The cascade screening has been proved as the most cost-effective method of identification and has a special importance for children [60], since a genetic defect can be identified before the atherosclerotic process, leading to the implementation of preventive measures and correct treatment.

Cascade screening might be life-saving, especially in young patients, since FH is a serious disease that requires early intervention, lifelong treatment and regular follow-up.

1.5. Aim of the project

Although the Portuguese FH Study has been implemented in Portugal in 1999 and more than 800 families have been enrolled [58], FH remains as an underdiagnosed disorder.

A clinical diagnosis is not enough to identify a FH patient, and only with molecular diagnosis is possible to identify the disease-causing variants. However, not all of identified variants are functionally assessed in order to understand their involvement in LDLR function. The performance of functional studies is imperative, providing a definite FH diagnosis.

In order to clarify the relationship between genotype and phenotype in FH, the aim of the present work is to perform the molecular diagnosis for 25 index cases with clinic criteria of FH.

Moreover, functional characterization of the 10 most common *LDLR* variants, which remain to date without functional studies, identified in Portuguese FH patients, also constitutes an aim, being the objective to determine the effect of these 10 *LDLR* variants in the cell surface expression, binding and uptake of the LDLR.

These results will allow the increase of knowledge about the functionality of LDLR variants, contributing to the elucidation of molecular basis of FH not only in Portugal, but worldwide, as some of the alterations under study were reported in several other countries.

The knowledge about genetic causes of FH and the relation between patients' genotype and phenotype, allows an accurate and definite diagnosis along with early personalized counselling and treatment, improving FH patients' prognosis and effectively reducing their cardiovascular risk.

Chapter 2

Materials and Methods

2.1. Molecular diagnosis

2.1.1. Patients Recruitment

All the patients were recruited for the Portuguese Familial Hypercholesterolemia Study, which protocol and database have been approved by the National Institute of Health Ethics Committee and the National Data Protection Commission, respectively.

During these past 15 years, patients with a clinical diagnosis of FH (Criteria in *Table 2.1.*, adapted from those of the Simon Broome Heart Research Trust [61]) have been recruited all over the country by clinicians from several specialties [58]. When a pathogenic variant is identified in a patient, the clinician is notified and asked to perform cascade screening in other relatives with and without a clinical diagnosis of FH for co-segregation analysis. Written informed consent was obtained from all participants before their inclusion in the study.

Table 2.1. – FH criteria adapted from "Simon Broome Heart Research trust"

Confirmed familial hypercholesterolemia is defined as: Index case: **Child under 16** with total cholesterol over 260 mg/dL (6.7 mmol/L) or LDL cholesterol over 155 mg/dl (4 mmol/L); Index case: **Adult** with total cholesterol over 290 mg/dl (7.5 mmol/L) or LDL cholesterol over 190 mg/dl (4.9 mmol/L), **and** Tendon xanthoma in the index case or relative (parents, children, grandparents, siblings, aunts or uncles),

or Genetic evidence of a variant in the *LDLR*, *APOB* or *PCSK9* genes.
Possible familial hypercholesterolemia is defined as: Index case: **Child under 16** with total cholesterol over 260 mg/dl (6.7 mmol/L) or LDL cholesterol over 155 mg/dl (4 mmol/L); Index case: **Adult** with total cholesterol over 290 mg/dl (7.5 mmol/L) or LDL cholesterol over 190 mg/dl (4.9 mmol/L), **and** Family history of myocardial infarction before the age of 50 in grandparents or aunts or uncles, or before the age of 60 in parents, siblings or children, and/or family history of elevated cholesterol levels (>290 mg/dL) in parents, siblings or children; or Total cholesterol (TC) over 290 mg/dl (7.5 mmol/L) in grandparents and/or aunts or uncles.

2.1.2. Blood samples collection

For each index case and respective relatives, fasting blood samples (7.5 mL in serum tube and 3 x 2.7 mL in EDTA tubes for adults; 5 mL in serum tube and 2 x 2.7 mL in EDTA tubes for children) were collected in order to perform DNA extraction. Moreover, 5 mL were collected in serum tubes to obtain serum, used in biochemical determination. For each sample a confidential identification number was assigned and all the information concerning the patients was registered in a confidential database, according to legal requirements. At maximum 48 h after the blood collection, the biochemical determination was performed.

2.1.3. Biochemical determination

The biochemical determination was executed by technicians of the Unidade Laboratorial Integrada at INSA. It included measurement of TC, HDL-C, LDL-C, triglycerides (TG), apoAI, apoB and Lp(a), performed in an autoanalyser Cobas Integra 400 Plus (Roche) by enzymatic, colorimetric and immunoturbidimetric methods.

2.1.4. Molecular biology techniques

The molecular analysis comprises 5 phases, as described in Introduction, section *1.4. Portuguese FH Study* [58].

2.1.4.1. Genomic DNA extraction

Genomic DNA was extracted from leucocytes in ~5 mL samples of peripheral blood, collected in EDTA tubes. For index cases, 10 mL were collected and DNA extraction was performed for both 5 mL tubes, in independent days, providing two different DNA samples for diagnosis confirmation. This extraction was performed as an adaption of the protocol described in [62]. The proportion of each reagent to mL of blood is disclosed in Appendix I, *Table A I.1.*

The blood was well homogenised and transferred to a 15mL falcon tube; it was added equal volume of TKM X-100 (low salt buffer containing 10 mM Tris-HCl pH 7.6, 10 mM KCl, 10 mM $MgCl_2$, 2 mM EDTA and 25 mL Triton-X 100/L) and mixed several times by inversion. Then IGEPAL was added, mixed until total solubilisation. The tubes were centrifuged – 10 min 2200 rpm at room temperature (RT) (centrifuge 5810 R, Eppendorf) and the supernatant was despised; the pellet was washed in TKM1 (TKM-X100 without the Triton-X 100) and centrifuged again for 10 min 1600 rpm at RT. The wash was repeated at least once, or twice when the pellet was still red. The pellet was resuspended with TKM2 (high salt buffer containing 10 mM Tris-HCl pH 7.6, 10 mM KCl, 10 mM MgCl2, 0.4 M NaCl and 2 mM EDTA) and then SDS 10% was added, followed by a 10 min incubation at 55ºC, for protein denaturation. All the content was transferred to a 2 mL Eppendorf tube, NaCl 5 M was added and mixed by inversion, for protein precipitation. The tubes were centrifuged for 20 min, 13200 rpm at RT, forming a dark protein pellet. All the supernatant was transferred to a clean falcon tube and absolute ethanol at RT was added and gently mixed by inversion, in order to denature and precipitate DNA. The DNA fibrils were removed with a loop and washed in a 70% ethanol solution at 4 ºC. When DNA was completely dry, it was resuspended in TE. DNA content was quantified in a spectrophotometer (NanoDrop 1000, Thermo Scientific), assessed by electrophoresis gel (2.1.4.3. Assessment of DNA fragments and PCR products) and stored in 1.5 mL tubes.

2.1.4.2. DNA amplification by Polymerase Chain Reaction (PCR)

Fragments containing parts of the *APOB* exons 26 and 29 of *APOB,* and the *LDLR* promoter and 18 exons, with respective flanking regions, were amplified by PCR. All the primers used, as well as distinct annealing temperatures for each one are written in Appendix I, *table A I.2.*

The PCR was performed with BIOTAQ™ DNA Polymerase kit (Bioline), according to the manufacturer's instructions: to a reaction tube 1 µl of dNTPs (100 mM dNTP Mix), 2.5 µl NH_4 buffer (10x, Bioline), 0.75 µl of Mg^{2+} (50 mM, Bioline), 1 µl of each primer (forward

and reverse) (10 ρmol/µl, Invitrogen), 1.25 U of BioTaq polymerase (Bioline) and bidistilled water was added up to a final volume of 24 µL. At last, 1 µl (100-200 ng) of genomic DNA was added. It was performed a replicate without DNA as a control for each exon amplification reaction.

The PCR reaction was held in a thermocycler (model 2720, Applied Biosystems), as follows: initial denaturation for 3 min at 95 ºC; 35 cycles of three steps: denaturation for 45 sec at 94 ºC, annealing for 30 sec at 57 ºC to 62ºC depending on the primer (Appendix I, *Table A.2.*), and elongation for 1 min at 72 ºC; and final extension for 30 min at 72 ºC. All PCR products were assessed by an agarose gel electrophoresis.

2.1.4.3. Assessment of DNA fragments and PCR products

An agarose gel was prepared in 100 mL of TBE buffer 1x (TBE 10x ultrapure, pH 8.4, 1.0 M Tris, 0.9 M boric acid, 0.01 M EDTA, Invitrogen). It was used a concentration (weight/volume) of 0.8% (w/v) for DNA qualitative assessment and of 1.5% (w/v) for PCR products visualization. SYBR safe (10000x concentrate in DMSO, Invitrogen) was added to the gel before polymerization. When polymerized (20 to 30 min), the samples (5 µL PCR product or DNA, gel loading die (Orange 6x, New Englands Biolabs) and bidistilled water to a final volume of 10 µL) were loaded, as well as a molecular weight marker (1 kB DNA Ladder, Boehringer Mannheim). The electrophoresis was performed for 40 min at 90 volt (Bio-Rad Power Pac 3000 equipment), in TBE 1x. The gel was visualized in a Safe Imager™ blue light transilluminator (Invitrogen).

2.1.4.4. Automated sequencing

Before Sanger sequencing, PCR products must be purified to remove the excess of primers and dNTPs. This was made by an enzymatic digestion using two hydrolytic enzymes; Exonuclease I and Shrimp Alkaline Phosphatase (SAP), combined in a commercial product named ExoStar (Illustra™ ExoStar™, GE lifesciences). To 2.5 µL of PCR product, 1 µL of ExoStar was added in a reaction tube and incubated for 15 min at 37 ºC (enzyme optimum temperature) and for 15 min at 80 ºC for enzyme inactivation. Purification products were stored at 4ºC until further use.

The sequencing reaction was prepared as follows: to a reaction tube, 2 ρmol of primer (Invitrogen), 1 µL of BigDye (Terminator Cycle Sequencing Ready reaction kit, Applied Biosystems) and bidistilled water was added up to a final volume of 9 µL. At last, 1 µL of purified DNA was added. The sequencing reaction was performed as follows: initial denaturation for 30 sec at 96 ºC; 25 cycles of three steps: denaturation for 10 sec at 96 ºC, annealing for 5 sec at 50 ºC and elongation for 4 min at 60 ºC. Sequencing products were stored at 4 ºC. Resulting products were sequenced by Unidade de Tecnologia e Informação (UTI – INSA) (3130xl Genetic Analyser, Applied Biosystems). The subsequent .AB sequence files were analyzed with Staden Package software (version 2.0). The reference sequence

NM_000527.4 for *LDLR* and NM_000384.2 for *APOB* was used and the novel variants were numbered according to the Human Genetic Variation Society (HGVS) guidelines, where +1 is the A of the ATG translation initiation codon of the coding DNA [63].

Whenever an alteration was found, all PCR, assessment and sequencing protocols were repeated in the second aliquot of index case, as well as made in its relatives, when samples were available, for confirmation and co-segregation studies.

2.1.4.5. Multiplex Ligation-dependent Probe Amplification (MLPA)

The search for large rearrangements (duplications or deletions) was performed by MLPA. This was made with SALSA® MLPA® kit (probemix P062-C2 LDLR, MRC-Holland, The Netherlands) according to the manufacturer's instructions. The MLPA reaction can be divided in five major steps: 1) DNA denaturation and hybridization of MLPA probes; 2) ligation reaction; 3) PCR reaction; 4) separation of amplification fragments by electrophoresis, which was performed by Unidade de Tecnologia e Informação (UTI – INSA) (3130xl Genetic Analyser, Applied Biosystems); and 5) data analysis, which was performed using the Coffalyser – MLPA analysis tool (developed at MRC-Holland, The Netherlands).

2.1.5. *In silico* analysis

Whenever a *LDLR* or *APOB* variant was found, the predicted effects of alterations were assessed using the following open access software: Polymorphism Phenotyping (PolyPhen-2) [64], Sorting Tolerant From Intolerant (SIFT) [65] and Mutationtaster [66] for prediction of single nucleotide substitutions. Briefly, SIFT [65] takes into account evolutionary conservation through the use of sequence alignments, while MutationTaster [66] and PolyPhen-2 [64] base their predictions in protein structure/function and evolutionary conservation, being that PolyPhen-2 also uses a prebuilt sequence alignment.

Human Splicing Finder (HSF) (http://www.umd.be/HSF3/HSF.html; [67]), the Splice Site Prediction by Neutral Network (NNSSP) (http://www.fruitfly.org/seq_tools/splice.html; [68]) and the FSPLICE (http://linux1.softberry.com/) tools were used for prediction of splicing defects. HSF [67] uses position-dependent logic, identifying exonic and intronic motifs, NNSSP [68] is based in neural networks combined a sequence similarity matrix with a local structural environment scoring scheme for predicting protein secondary structure and FSPLICE (http://linux1.softberry.com/) bases its predictions on weight matrices model, which consider the importance of the presence of a determinate nucleotide in a specific position.

Mutation Taster also predicts a phyloP score. PhyloP score is a measurement of evolutionary conservation, thus the higher its score is, the stronger is the evolutionary conservation for a specific nucleotide.

A variant was considered probably pathogenic when all three software tools had a prediction of pathogenic (probably damaging, deleterious or disease causing), variant of unknown significance (VUS) if predictions were contradictory and neutral when tools predicted it to be benign (benign, tolerated or polymorphism). Regarding splice-site analysis tools, a variant was considered pathogenic if there was a deletion of the actual splice-site or the addition of a new one.

2.2. Production of *LDLR* gene variants

2.2.1. Site-directed mutagenesis

In order to express all 10 different alterations in the LDLR gene, individual point mutations were introduced into the human LDLR cDNA, previously subcloned in the mammalian expression vector pcDNA3, under the control of a SV40 promoter (pcDNA3_LDLR, kindly offered by César Martín) (*Appendix 2, Figure A II.1.*). This was performed by oligonucleotide site-directed mutagenesis with the Quik-Change XL mutagenesis kit (Agilent Technologies) according to the manufacturer's instructions. This procedure is based on a PCR reaction that uses specific complementary mutagenic primers of the fragment of the interest (insert – complete LDLR cDNA) (*Table 2.2.*).

Each mutagenesis reaction tube was composed of 5 µL of reaction buffer (10x, Agilent Technologies), 125 ng of each pair of mutagenic primers (forward and reverse), 1 µL of dNTPs, 1 µL of PfuUltra High Fidelity DNA polymerase (2.5 U/µl), 50-100 ng of double-stranded plasmid DNA (pcDNA3_LDLR) and bidistilled water up to 50 µL. The PCR program was performed as follows: initial denaturation for 10 min at 95 ºC; 12 cycles of three steps: denaturation for 30 sec at 90 ºC, annealing for 1 min at 55 ºC and extension for 12 min at 68 ºC. Completed the program, 10 µL of the products were visualized in an 1% agarose gel (see Annex III, figure A III.1.), as well as 50 ng of the original plasmid (the same amount of DNA of the mutagenesis reaction) properly mixed with DNA loading dye and bidistilled water up to a final volume of 20 µL. Next, the mutagenesis products were treated with the restriction enzyme DpnI (provided with the kit) for 1 h at 37 ºC, in order to digest the parental DNA that does not contain the desired alteration. This strategy is based in the fact that plasmids grown in almost all E. coli strains are dam methylated and therefore are susceptible to DpnI specific digestion of methylated and hemimethylated DNA.

Table 2.2. LDLR variants and negative control (for LDLR expression) under functional assessment and respective mutagenic primers.

Variant	Exon	5' - 3' Sequence	Primers
c.1775 G>A, p.(Gly592Glu)	Exon 12	F: GCATCGATGTCAACGAGGGCAACCGGAAGAC	MB480
		R: GTCTTCCGGTTGCCCTCGTTGACATCGATGC	MB481
c.618_638del p.(Gly207_Ser213del)	Exon 4	F: GTTCCACTGCCTAAG///CTGGCGCTGTGATGG	MB420
		R: CCATCACAGCGCCAG///CTTAGGCAGTGGAAC	MB421
c.2177 C>T p.(Thr726Ile)	Exon 15	F: CCACCCAGGAGACATCCATCGTCAGGCTAAAGGTCAG	MB470
		R: CTGACCTTTAGCCTGACGATGGATGTCTCCTGGGTGG	MB471
c.551G>A p.(Cys184Tyr)	Exon 4	F: GTGGCCGCAGCGCTATAGGGGTCTTTACG	MB418
		R: CGTAAAGACCCCTATAGCGCTGCGGCCAC	MB419
c.1816G>T p.(Ala606Ser)	Exon 12	F: AGGATGAAAAGAGGCTGTCCCACCCCTTCTCCTTG	MB430
		R: CAAGGAGAAGGGGTGGGACAGCCTCTTTTCATCCT	MB431
c.1876G>A p.(Glu626Lys)	Exon 13	F: TTTTGGACAGATATCATCAACAAAGCCATTTTCAGTGCCAACC	MB432
		R: GGTTGGCACTGAAAATGGCTTTGTTGATGATATCTGTCCAAAA	MB433
c.1802A>T, p.(Aps601Val)	Exon 12	F: GAAGACCATCTTGGAGGTTGAAAAGAGGCTGGCCC	MB478
		R: GGGCCAGCCTCTTTTCAACCTCCAAGATGGTCTTC	MB479
c.631C>G p.(His211Asp)	Exon 4	F: TGGCGAGTGCATCGACTCCAGCTGGCG	MB422
		R: CGCCAGCTGGAGTCGATGCACTCGCCA	MB423
c.1966C>A p.(His656Asn)	Exon 14	F: GGATATGGTTCTCTTCAACAACCTCACCCAGCC	MB468
		R: GGCTGGGTGAGGTTGTTGAAGAGAACCATATCC	MB469
c.661G>T p.(Asp221Tyr)	Exon 4	F: CTGTGATGGTGGCCCCTACTGCAAGGACAAATC	MB424
		R: GATTTGTCCTTGCAGTAGGGGCCACCATCACAG	MB425
c.1633G>T p.(Gly545Trp)	Exon 11	F: CTGCCAAGATCAAGAAATGGGGCCTGAATGGTGTG	MB472
		R: CACACCATTCAGGCCCCATTTCTTGATCTTGGCAG	MB473

/// Represents the 21bp deletion (5' TGGCGAGTGCATCCACTCCAG 3')

2.2.2. Bacteria transformation; Plasmid DNA extraction and purification

Competent cells (NZY5α Competent Cells, NZYTech) were transformed with DpnI digestion product: to a volume of 50 µL of bacteria, 5 µL of mutated DNA were added in a reaction tube, placed in ice to allow the DNA adsorption to the bacteria wall. After 1 h, bacteria underwent a heat shock: 40 sec at 42 ºC and then immediately placed on ice for 2 min, in order to create pores in the membrane and allow the entrance of plasmid DNA. A volume of 250 µL of Luria Broth (LB) medium was added to the tube and placed at 37 ºC for 1 h with constant agitation for bacteria to generate the antibiotic resistance encoded in the plasmid DNA. Each reaction was plated in LB-Agar (Lennox L. Agar, Invitrogen, 32% [w/v]) with 100 µg/mL ampicillin (Sigma). After 16 h, the grown clones were selected, resuspended in 3 mL of LB medium with ampicillin (100 µg/mL) in Falcon tubes and grown for 16 h at 37 ºC with constant agitation (220 rpm). Bacteria were lysed and the plasmid DNA was extracted and purified with JETQUICK Plasmid Miniprep Spin kit (Genomed),

according to manufacturer's instructions. The DNA concentration was determined by absorbance lecture in a spectrophotometer (NanoDrop 1000, Thermo Scientific).

2.2.3. Confirmation of variant and insert of interest

The success of the mutagenesis protocol in each clone was confirmed by direct sequencing (see 2.1.4.4. Automated Sequencing) of the LDLR fragment. All primers used for sequencing are disclosed in *Appendix I, Table A I.3*.

When the presence of the desired variant was confirmed, the insert of interest (complete LDLR cDNA), as well as the adjacent regions in the vector were sequenced, in order to verify if the introduced alteration was the only present in the sequence. Resulting sequences were analysed with Chromas Lite software (version 2.1.1.).

2.2.4. Recloning

A recloning step was performed, in which the insert of interest (LDLR variant) was withdrawn and placed into a clean vector (pcDNA3), which was not subjected to the mutagenesis process. A simplified scheme is represented in Appendix II, Figure A II.2.

In a reaction tube, 1 µg of the clean construct pcDNA3_LDLR (not subjected to site-directed mutagenesis) was digested with 0.5 µL of each enzyme (XbaI and KpnI, 10 U/µL, Thermo Scientific), 5 µL of tango buffer (10x Thermo Scientific) and bidistilled water up to a final volume of 50 µL. Independently, 2 µg of each mutated plasmids were digested with 1 µL of each enzyme (XbaI and KpnI), 5 µL of tango buffer (10x) and bidistilled water up to 50 µL. A control without enzymes was made for each plasmid DNA, in the same conditions. The reaction tubes were placed for 2 h at 37 °C.

All enzymatic digestions were confirmed in a 0.8% agarose gel (*Annex III, figure A III.1*.). Each sample showed two bands, corresponding to the vector pcDNA3 and to the insert of interest LDLR. The band of the clean vector was isolated from the agarose gel, as well as the band of the mutated insert of interest. Both were purified with the JETQUICK Gel Extraction Spin kit (Genomed), according to the manufacturer's instructions. The resulting DNA concentrations were measured.

The ligation reaction between the clean vector (dephosphorylated) and the mutated insert of interest was performed with T4 ligase (Rapid DNA Dephos & Ligation kit, Sigma Aldrich), according to the manufacturer's instructions, using a molar ratio of 1:7. This process leads to a construct composed of a pcDNA3 vector that did not undergo site-directed mutagenesis and a LDLR with the desired variant inserted. The resulting DNA was used to transform bacteria as described in *2.2.2. Bacteria Transformation, Plasmid DNA extraction and purification*.

The success of ligation reaction was confirmed by sequencing of the restriction and the variant sites. Hence, the plasmids are ready to be used in functional studies in cells.

2.2.4.1. Colony PCR

When bacteria were transformed with plasmids, theoretically carrying the deletion c.618_638del p.(Gly207_Ser213del), a lot of colonies were observed. For confirmation of the presence of this deletion, a colony PCR was performed (*Annex III, figure A III.2.*).

After 16 h of incubation at 37 ºC, as described in section 2.2.2. Bacteria transformation, Plasmid DNA extraction and purification, all colonies were selected and washed in 10 µL of milli-Q water. The colony PCR was performed with 2.5 µL of the bacterial solution mentioned above, primers MB10 and MB11 (primers' sequences in Appendix I, *Table A I.3.)* and GoTaq® DNA polymerase (Promega) according to the manufacturer's instructions. The PCR reaction was held as follows: initial denaturation for 10 min at 98 ºC; 35 cycles of three steps: denaturation for 30 sec at 95 ºC, annealing for 30 sec at 58 ºC, and elongation for 90 sec at 72 ºC; and final extension for 5 min at 72 ºC. Resulting PCR products were visualized in a 1.5 % agarose gel.

For colonies which amplified fragments seemed, in the agarose gel, to present the deletion, the bacterial solution was used to grown a liquid culture for 16 h at 37 ºC in LB medium and plasmid DNA extraction was performed, as described in 2.2.2. Bacteria transformation; Plasmid DNA extraction and purification. The presence of deletion was confirmed, in these minipreps, by automated sequencing.

All steps concerning 2.2. Production of *LDLR* gene variants are schematically represented in Appendix II, *figure A II.2*.

2.3. Functional studies

In order to evaluate the impact of variants under study in LDLR function, functional assessment was performed as previously described [69].

2.3.1. Expression of LDLR proteins *in vitro*

2.3.1.1. Cell culture

LDLR-deficient Chinese hamster ovary (CHO) cell line ldlA7 (CHO-ldlA7), without endogenous expression of *LDLR,* was cultured in 75 mm flasks with Ham's F-12 medium (Nutrient Mixture F-12 HAM, Sigma Aldrich) supplemented with 10% of fetal bovine plasma (FBS) (Thermo Scientific), 2 mM L-glutamine, 100 units/mL penicillin and 100 µg/mL streptomycin (PenStrep Glutamine 100x, Life Technologies), at 37 °C and 5% CO_2.

When cells presented the desired confluence, 50000 cells/well were plated in 24 well culture plates in 1 mL of opti-MEM for 24 h. Cells were transfected with plasmids carrying

the individual *LDLR* variants, as well as with a wt *LDLR* plasmid and two control plasmids, for surface expression, binding and uptake assessment.

2.3.1.2. Transfection

Transfection was performed using Lipofectamine LTX and Plus Reagent (Invitrogen), according to the manufacturer's instructions, as follows. For each well two mixes were prepared: the first mix with 1 µg of plasmid DNA, 1 µL of Plus Reagent and 25 µL Opti-MEM (Reduced Plasma Medium, Life Technologies); the second mix was constituted by 2 µL Lipofectamine and 25 µL of Opti-MEM. After 5 min (to avoid Lipofectamine cytotoxic effects), both were mixed and then awaited for 30 min, allowing the formation of Lipofectamine-DNA complexes.

From each well, 500 µL of opti-MEM were withdrawn and 50 µL of transfection mix was added to the well (to a final concentration of 10%) and incubated for 6 h at 37 ºC. In order to recuperate the transfected cells, 500 µL of Opti-MEM - 5% FBS were added to each well and cells were maintained in culture for 24 h to achieve maximal LDLR expression.

2.3.2. Flow cytometry

LDLR expression at cell surface, binding and uptake were then assessed by flow cytometry.

Flow cytometry is a technique in which multiple physical characteristics of cells can be measured and analysed, as the cells flow in a fluid stream through a laser. When these cells pass through the laser intercept, they scatter laser light and emit fluorescence according to their properties.

The argon ion laser is commonly used in flow cytometry because the 488-nm light that it emits excites more than one fluorochrome. Among these, is Alexa Fluor 488 Dye, which is the fluorophore used for immunodetection in expression assays (see section *2.3.2.2. Expression assessment by flow cytometry*). Furthermore, FITC is also excited at this wavelength. For this reason, the LDL labelling for binding and uptake assays was performed with LDL labelled with this compound (see section *2.3.2.3. Binding and uptake assessment by flow cytometry*).

2.3.2.1. Lipoprotein labelling with FITC

LDL was purified from blood samples following a procedure previously described [70]. Lipoproteins were separated following a density gradient created with KBr (Scharlau). For this, 4.84 g of KBr were dissolved in 4 mL of plasma and PBS was gently added in order to create a difference of densities, resulting in a final average density of 1.21 g/ml. Tubes were centrifuged for 16 h at 4 ºC, 35400 rpm in a Centrikon T-2190 ultracentrifugue (Kontron).

The fraction corresponding to LDL was collected and labelled with FITC. LDL (approximately with a density of 1 mg/mL) was passed through a column previously calibrated with 0.1 M $NaHCO_3$ (pH 9, which favours FITC-Lysine bonding). Then, it was gently mixed with 10 µL/mL FITC (2 mg/mL in dimethyl sulfoxide), by slow rocking at room temperature for 2 h. The unreacted dye was removed by gel filtration on a Sephadex G-25 column equilibrated with PBS EDTA-free buffer. All fractions were assayed for protein content with bovine plasma albumin (Pierce BCA protein assay, Pierce).

2.3.2.2. Expression assessment by flow cytometry

To determine cell surface expression of *LDLR* by fluorescence activated cell sorting (FACS), transfected cells grown during 24 h (as well as controls) were washed three times with filtered PBS 1% BSA, fixed with 200 µL of paraformaldehyde 4% for 10 min at room temperature and washed again three times. Fixed cells were incubated with the primary antibody, a mouse monoclonal anti-LDLR antibody (1:100; 5 mg/L; Progen Biotechnik GmbH, Cat. No. 61087), overnight at 4 ºC with gentle agitation. This monoclonal antibody recognizes an epitope in the region of repeat 1 of the ligand binding domain. Cells were washed three times with PBS - 1% BSA and incubated with a goat anti-mouse IgG Alexa Fluor 488-conjugated secondary antibody (1:200, Molecular Probes, Cat. No. A-11001) (polyclonal) for 1 h. Cells were once again washed three times with PBS - 1% BSA and remained in 300 µL of PBS. Cells were gently scraped and fluorescence intensities were measured by Flow Cytometry, in a FACScalibur™ flow cytometer.

2.3.2.3. Binding and uptake assessment by flow cytometry

CHO-ldlΔ7 transfected cells grown for 24 h were incubated for 4 h with 20 µg/mL FITC-LDL.

For binding assessment, the incubation was performed at 4 ºC, since LDL is not internalized at this temperature, remaining bound to the LDLR at the cell surface. Cells were washed three times, fixed in 200 µL of paraformaldehyde 4% for 10 min, and washed again three times. In order to keep the receptor's structure, responsible for LDL binding, all washes were performed with a buffer with calcium (150 mM NaCl, 2 mM $CaCl_2$, 20 mM Hepes, pH 7.5) plus 1% BSA. Cells were resuspended in 300 µL of PBS, being after gently scraped and fluorescence intensities were measured by flow cytometry.

To assess the uptake, the incubation was performed at 37 ºC. After 4 h, cells were washed three times with PBS - 1% BSA, fixed in 200 µL of paraformaldehyde 4% for 10 min, washed again three times with PBS - 1% BSA and then awaited in 200 µL of PBS- 1% BSA. Cells were gently scraped and Trypan blue solution (Sigma Aldrich, Steinheim, Germany) was added to a final concentration of 0.2%, directly to the samples and measured by flow cytometry. Trypan blue solution quenches external fluorescence and allows the distinction between internalized and surface-adherent FITC-LDL particles. Only the

remaining fluorescence of the LDL particles inside cells is measured, since it is not affected by the external quencher.

2.3.2.4. Measurements by Flow Cytometry

Analysing samples in a flow cytometer allows the measurement of several properties of a cell at the same time, while passing suspending cells in a stream of fluid by lasers. Deviations of these lasers performed by cells are detected, giving information about cell composition.

Forward-scattered light (FSC) is proportional to cell-surface area or size, whereas side-scattered light (SSC) is proportional to cell granularity or internal complexity [71]. Applying gates to these parameters, selecting an interval for size and complexity of our cells, makes possible the selection of a desired population.

In order to analyse the results obtained, four controls were established. Variant Ex3_4del was used as a positive control for cell surface expression and negative control for binding and uptake at the same time, in order to guarantee that *LDLR* can be fully expressed, being its function defected at later steps of the pathway. Variant p.(Gly545Trp) was used as a negative control for cell surface expression, being consequently a negative control also for binding and uptake, since an unexpressed variant will never be able to correctly perform its function. Furthermore, a wt plasmid was also used, as well as a control of untransfected cells.

For each sample, fluorescence of 10000 events were acquired for data analysis. All measurements have been performed in triplicate. All analysis were performed using Flowing software, which is a university research project, not a commercial software (Cell Imaging Core, Turku Centre for Biotechnology, Finland).

When analysing these data, the desired population is selected, excluding dead cells resultant from transfection and physic scrape. It is possible to obtain a graphic representation, as the one in *figure 2.1.* It is important to refer that this representation urges as an example, which could have resulted from expression, binding or uptake assessment.

Although this kind of graphical representations is quite informative, the final purpose is to obtain the percentage of expression, binding and uptake in order to infer about the receptor activity in percentage. For this, two areas are considered (*figure 2.1.*).

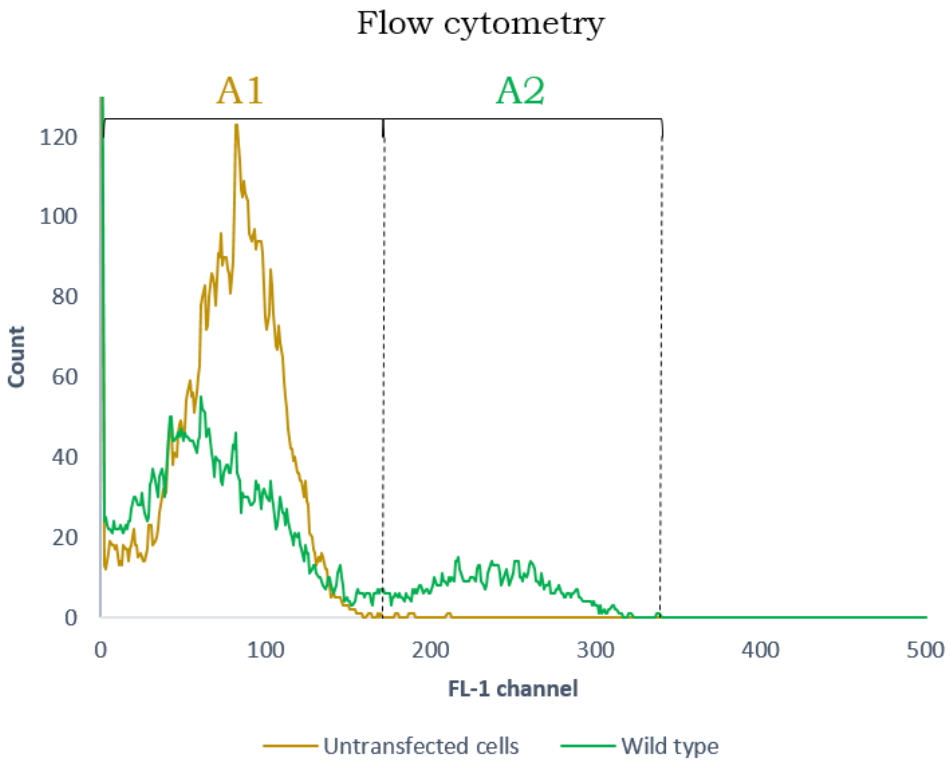


Figure 2.1. Count of events in function of fluorescence measured through the FL-1 channel. *First area (A1) represents peaks corresponding to basal fluorescence of cell membrane, while area A2 corresponds to additional fluorescence, which occurs due to the immunodetection of LDLR at the membrane surface by antibodies linked to the fluorophore Alexa-488.*

In *figure 2.1.* is possible to observe a yellow line, corresponding to the measurement of untransfected cells, and a green line, corresponding to the measurement of cells transfected with a wt plasmid.

The yellow line only presents one peak, corresponding to the basal fluorescence, delimited in the A1 area. The green line presents two peaks: the first one, also included as well into A1, is the measurement of basal fluorescence of untransfected cells among the population measured. The second peak, at higher values of fluorescence and delimited into area A2, corresponds to the fluorescence measured in cells transfected with the wt plasmid. This peak could be representative of a wt expression, due to immunodetection of LDLR at cell superficie by antibodies bound to a fluorochrome (Alexa-488); or this peak could be representative of wt binding or uptake, due to the fluorescence of FITC-labelled LDL.

After selecting these areas, values corresponding to the geometric mean of each area were withdrawn for all analysed samples. Geometric means take into account the fluorescence values obtained for each area, but considering the number of events here measured. Lower fluorescence values are often observed for A1, as it is constituted by a big peak, indicating a large number of cells, but low fluorescent values, corresponding to the cell basal fluorescence. Contrarily, higher values are often observed for A2, as it is

observable a smaller peak, correspondent to the transfected population of cells, which emits a higher fluorescence, due to the presence of the LDLR.

For control populations (yellow line in *figure 2.1*.), a second peak is not supposed to be observed, so geometric mean of A2 will not be considered. For the remaining populations, the geometric mean of A2 was selected and then subtracted the geometric mean of A2 area corresponding to a control population. This way, basal fluorescence will be subtracted from the observed fluorescence.

All fluorescence values were normalized to the values of the fluorescence concerning the cells transfected with the wt plasmid, i.e. geometric mean of A2 area of the line corresponding to wt, and a graphic representation was drew.

Statistical analysis was performed using Student's t-test and changes were considered to be significant when p-value was below 0.01.

Chapter 3

Results

3.1. Molecular diagnosis

A total of 25 unrelated individuals were studied. The pediatric group is constituted by 9 index cases and the adult group by the remaining 16. All of them were admitted to the study since they fulfilled clinical criteria according to the SB criteria (Materials and Methods, *Table 2.1.*). A total of 9 relatives, affected and unaffected, were also admitted. Clinical and biochemical characterization, as well as relevant information as age, cardiovascular events and medication is presented in *table 3.1*.

Analysis of the promotor, 18 exons of *LDLR* and adjacent regions, as well as part of exons 26 and 29 of *APOB*, for all patients, led to the finding of putative disease-causing variants, as shown in *table 3.2.* Electropherograms of variants found are shown in *Appendix III – Sequence analysis*.

In this study, no homozygous patients were identified. From 25 index cases, 12 were identified with a heterozygous variant *(table 3.2.).* Although only 10 variants were considered at least probably pathogenic. There was considerably dispersion of values and it was not possible to distinguish individually between patients with and without a variant based only on lipid profile, revealing once again the importance of the molecular diagnosis.

Table 3.1. Clinical and biochemical characterization of studied index cases.

	Patient	Sex	Age	Biochemical Values								CV events	Family history of CHD	Family history of high cholesterol	Medication
				TC (mg/dL)	**LDL-C** (mg/dL)	**HDL-C** (mg/dL)	**TG** (mg/dL)	**Lp(a)** (mg/dL)	**ApoA1** (mg/dL)	**ApoB** (mg/dL)	**ApoB/ ApoA1**				
Pediatric group	1	F	10	213	172	56	36	236	161	102	0.63	-	No	Yess	-
	2	F	12	389	298	53	127	23	141	193	1.4	-	No	Yes	-
	3	M	9	314	246	56	72	27*	163*	111*	0.7*	-	No	Yes	Statins
	4	F	8	373	292	65	79	2*	153*	101*	0.7*	-	Yes	Yes	Statins
	5	F	3	222	157	56	42	36,3	137	104	0.8	-	No	Yes	-
	6	M	6	320	223	83	68	50,5	205	109	0.5	-	No	Yes	-
	7	F	14	232	158	52	106	7,8	154	99	0.64	-	Yes	Yes	-
	8	M	8	249	195	34*	85	6,6*	119*	91*	0.76*	-	No	Yes	Statins
	9	F	6	311	237	60	53	15,6	136	160	1.2	-	No	Yes	-
Adult group	10	M	44	323	239	52*	178	188*	152*	111*	0.73*	-	No	Yes	Statins
	11	F	52	302	199	49	73	2*	154*	70*	0.45*	-	No	Yes	Statins + Ezetimibe
	12	F	71	400	114	59	109	26*	156*	74*	0.47*	Angina, (61); MI, CABG, (62)	No	Yes	Statins
	13	M	41	151	82	48	49	117*	135*	70*	0.52*	MI, PTCA (37)	Yes	Yes	Statins
	14	M	37	297	202	32	314	10*	99*	80*	0.8*	MI, PTCA (37)	Yes	Yes	Statins
	15	M	29	415	341	38	57	12	134	218	1.6	-	Yes	Yes	-
	16	M	45	311	212	73	80	5	176	143	0.8	-	No	Yes	-
	17	M	55	237*	167*	58*	64*	47,6*	142*	125*	0.9*	MI, PTCA (51)	No	Yes	Statins + Ezetimibe + Fibrates
	18	F	72	332	222	95	108	2*	208*	141*	0.7*	-	Yes	Yes	Statins + Ezetimibe
	19	M	34	321	206	84	114	11,3	209	127	0.6	-	No	Yes	-
	20	M	34	273	189	59	127	93*	143*	69*	0.48*	MI (34)	No	Yes	Statins
	21	F	53	271	213	54	82	1.3*	141*	141*	1*	Angine (52)	Yes	Yes	Statins
	22	F	45	331	245	65	107	37.8	153	123	0.8	-	Yes	Yes	-
	23	F	62	288	183	75	152	1,5*	211*	93*	0.4*	-	No	Yes	Statins
	24	F	75	232*	151*	64*	140*	109.9*	183*	104*	0.6*	-	No	Yes	Statins
	25	F	51	274	190	62	108	2.1*	149*	125*	0.8*	-	No	Yes	Ezetimibe

TC, total cholesterol; LDL-C, Low-density lipoprotein cholesterol; HDL-C, High-density lipoprotein cholesterol; TG, triglyceride; Lp(a), Liproprotein (a); ApoAI, Apolipoprotein AI; ApoB, Apolipoprotein B; MI, Myocardial infarction; CABG, Coronary artery bypass grafting; PTCA, Percutaneous transluminal coronary angioplasty.

* Values on cholesterol-lowering medication as stated.

Table 3.2. *Variants in LDLR and APOB genes identified in studied index cases.*

Patient	Alteration				First described by	Functional studies by	Coseg	Portuguese Normolipidemic panel	*In silico* Prediction							Classification
	Localization	cDNA	protein	Domain					Mutation Taster	Polyphen 2	SIFT	PhyloP	Human Splicing Finder	NNSSP	FSPLICE	
2	*LDLR* Exon 11	c.1633G>T	p.(Gly545Trp)	EGF precursor homology	[72]	[73]	2/2;0/1	0/208 non-FH alleles	Disease-causing	Probably Damaging	Damaging	1.662	100 (11A), 100 (11D)	100 (11A), 100 (11D)	100 (11A), 100 (11D)	Probably pathogenic
3	*LDLR* Intron 7	c.1060+1G>A	p.Gly314Aspfs*507 (skipping of exon 7)	EGF precursor homology	[72]	[74]	2/2;0/0	0/150 non-FH alleles	Polymorphism	NA	NA	0.21	100 (7A), 69 (7D)	100 (7A), 0 (7D)	100 (7A), 0 (7D)	Probably pathogenic
4	*LDLR* Exon 4	c.670G>A	p.(Asp224Asn)	Ligand-binding	[15]	[15]	2/2;0/0	0/190 non-FH alleles	Disease-causing	Probably Damaging	Damaging	5.864	100 (4A), 100 (4D)	100 (4A), 100 (4D)	100 (4A), 100 (4D)	Probably pathogenic
5, 6	*APOB* Exon 26	c.10580G>A	p.(Arg3527Gln)	Domain 6	[75]	[76]	3/3;0/3	0/190 non-FH alleles	Disease-causing	Probably Damaging	Damaging	4.569	100 (26A), 100 (26D)	100 (26A), 100 (26D)	100 (26A), 100 (26D)	Probably pathogenic
8	*LDLR* Exon 4	c.589T>C	p.(Cys197)Arg	Ligand-binding	[77]	[77]	2/2;0/1	0/190 non-FH alleles	Disease-causing	Probably Damaging	Damaging	1.777	100 (4A), 100 (4D)	100 (4A), 100 (4D)	100 (4A), 100 (4D)	Probably pathogenic
9	*LDLR* Intron 3	c.313+1G>A	p.Leu64_Pro105delinsSer / p.Pro105_Ala860delinsArgLysCysGlyProAlaPheAlalleGluProIle (two different transcripts)	EGF precursor homology	[78]	[78][79][80]	1/1;0/0	0/208 non-FH alleles	Polymorphism	NA	NA	0.152	100 (3A), 67 (3D)	100 (3A), 0 (3D)	100 (3A), 0 (3D)	Probably pathogenic
15	*LDLR* Exon 3	c.301G>A	p.(Glu101Lys)	Ligand-binding	[81]	[15]	1/1;0/0	0/208 non-FH alleles	Disease-causing	Probably Damaging	Damaging	4.232	100 (3A), 100 (3D)	100 (3A), 100 (3D)	100 (3A), 100 (3D)	Probably pathogenic
17	*LDLR* Exon 4	c.618_638del	p.(Gly207_Ser213del)	Ligand-binding	[72]	-	1/1;0/0	0/190 non-FH alleles	Disease-causing	NA	Deleterious	NA	100 (4A), 100 (4D)	100 (4A), 100 (4D)	100 (4A), 100 (4D)	Probably pathogenic
20	*LDLR* Intron 8	c.1186+56_1186+64del	p.?	EGF precursor homology	novel	-	1/1;0/0	0/150 non-FH alleles	Polymorphism	NA	NA	NA	100 (8A), 100 (8D)	100 (8A), 100 (8D)	100 (8A), 100 (8D)	Neutral
22	*LDLR* Exon 10	c.1585G>C	p.(Gly529Arg)	EGF precursor homology	novel	-	1/1;0/0	0/208 non-FH alleles	Disease-causing	Probably Damaging	Damaging	3.016	100 (10A), 100 (10D)	100 (10A), 104 (10D)	100 (10A), 138 (10D)	Probably pathogenic
23	*LDLR* Exon 15	c.2177C>T	p.(Thr726Ile)	O-linked sugars	[15]	-	1/1;0/0	0/200 non-FH alleles	Polymorphism	Benign	Tolerated	1.027	100 (15A), 100 (15D)	100 (15A), 100 (15D)	100 (15A), 100 (15D)	Neutral

Coseg, co-segregation in studied families, variant carriers/total affected; variant carriers/total non-affected; NA, not applicable. *In silico* prediction: SIFT, sort intolerant from tolerant; NNSSP, Nearest-neighbor Secondary Structure prediction; FSPLICE, find splice sites in genomic DNA; A, Acceptor site; D, Donor site.

These variants were mainly identified in the pediatric group, in 7/9 individuals. All identified variants were previously reported as pathogenic and present functional studies *(table 3.2.),* giving a positive rate of 78% in the children group. In adults, the definitive diagnosis was only possible for one case, as for the other 4 variants identified in this group, pathogenicity is unknown and functional studies need to be performed. However, according to *in silico* classification, only 1 of these 3 variants is probably pathogenic.

Among 11 variants identified, one missense *APOB* variant, already proved to be pathogenic, was found in two cases (patients 5 and 6). The remaining 10 were found in *LDLR:* 7 missense and 3 splicing variants, as well as a deletion of 21bp. Among these, 8 have been previously reported (only 6 of them with functional studies) and 2 are novel (*figure 3.1.*).

Several variants found were first described in other populations (see *table 3.2.*), with exception of variants c.618_638del p.(Gly207_Ser213del), c.1060+1 G>A and c.1633G>T, which were first described in Portugal, two of them already functionally characterized.

Concerning co-segregation of these variants with the phenotype, it was observed that, in families in which samples of relatives were available to study, all co-segregated. Furthermore, none of them was found in the Portuguese normolipidemic panel.

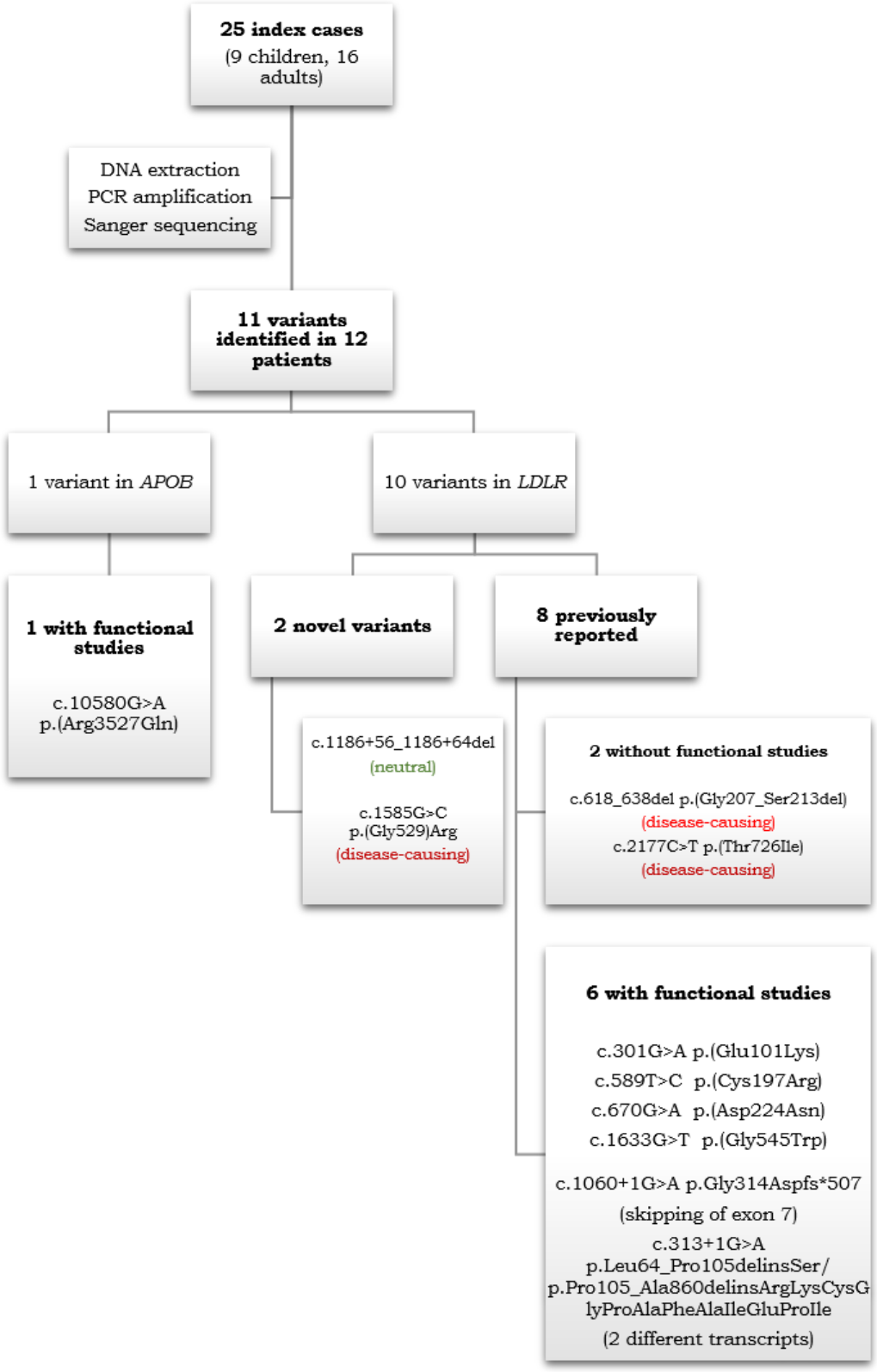


Figure 3.1. Schematic representation of Molecular diagnosis results. *25 index cases were assessed and 11 variants were found in 12 patients. Among these, 1 was found in APOB and had already been proved to be pathogenic. The remaining 10 were identified in LDLR: 2 were novel and 8 have been previously reported. Among all identified variants, 4 still need functional studies in order to prove their pathogenicity, although 3 are predicted to be pathogenic by in silico tools.*

MLPA analysis

Included in the molecular diagnosis, the assessment of large rearrangements by MLPA was performed. Data analysis revealed no alterations of this type in the group of patients under study (*figure 3.2.*).

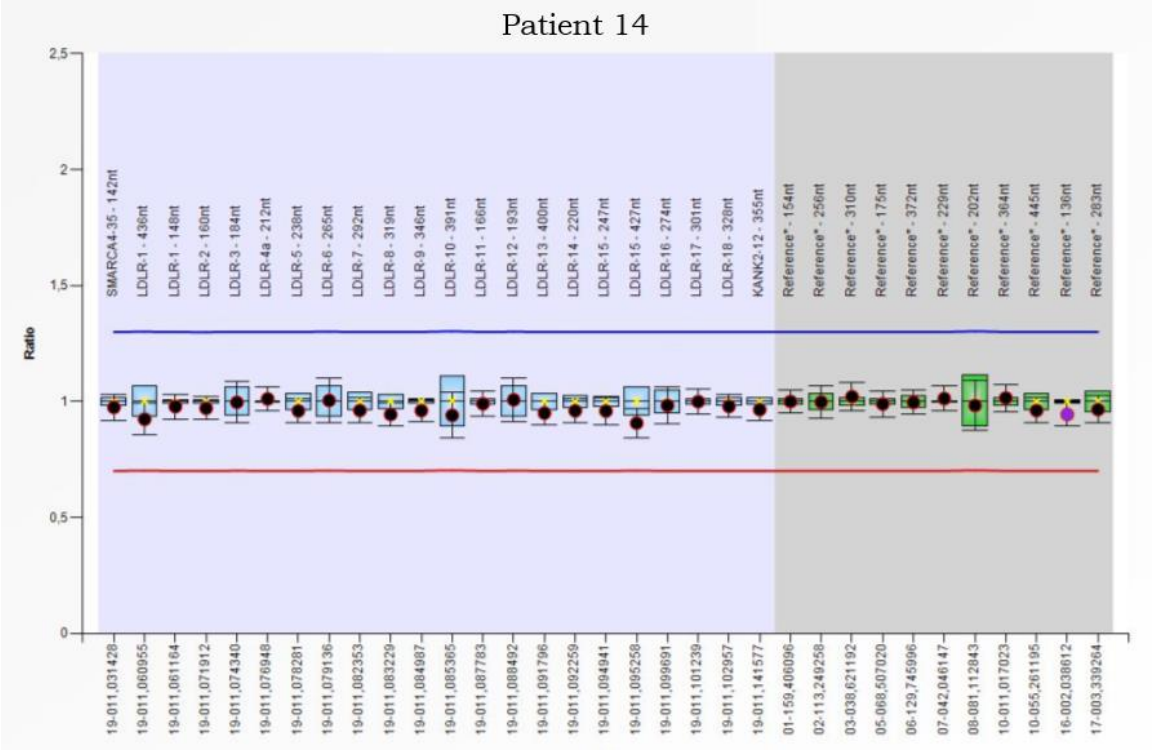


Figure 3.2. MLPA results *(example of patient 14). Values in the x-axis represent the location in the chromosome 19 of amplified probes for all exons, as well as for reference probes. Results are presented as the ratio between control and patient 14 probes. The ratio of all probes represented is approximately 1, indicating no large rearrangements in this analysis. This output was obtained with Coffalyser – MLPA analysis tool (developed at MRC-Holland, The Netherlands), as described in Materials and Methods, section 2.1.4.5.*

In silico prediction

Although 7 of the 11 variants found in these patients had already functional studies proving their pathogenicity, *in silico* prediction was also performed to all of them (*table 3.2.)*, where only two were predicted to be neutral (c.2177C>T p.Thr726Ile and c.1186+56_1186+64del).

Effects on splicing were also predicted for all variants. Intronic variants c.313+1G>A and c.1060+1G>A were predicted to be probably pathogenic and c.1186+56_1186+64del neutral. All missense variants were classified as neutral by these tools.

Moreover, the majority of the variants presented a high PhyloP value, indicating their rich evolutionary conservation, suggesting that as conserved nucleotides, the amino acids

codified by them must play important roles in protein structure and function. Nevertheless, variants c.589T>C p.(Cys197Arg), c.1633G>T p.(Gly545Trp) and c.2177C>T p.(Thr726Ile) constituted an exception, presenting low PhyloP values, associated with poorly preserved amino acids. Concerning their location, among 10 variants found in *LDLR*, 4 occurred in the ligand-binding domain, 5 in the EGF precursor homology and 1 in O-linked sugars domain.

Regarding *in silico* analysis, all variants with functional studies proving their pathogenicity were predicted, by *in silico* tools, to be probably pathogenic, revealing a 100% (7/7) of variants correctly predicted.

Furthermore, if all variants predicted to be pathogenic are indeed pathogenic, this would give a positive identification rate of 10/25 (40%) in this study.

3.2. Production of *LDLR* variants

Ten most common *LDLR* variants found in Portuguese FH Study, still without functional studies to date *(figure 3.3.)* were assessed, as well a control for *LDLR* activity, c.1633G>T (p.(Gly545Trp)), which has already been proved as pathogenic (for references, *table 3.3.)*.

Altogether, these ten variants were previously identified in a total of 97 patients, who remain with an uncertain FH diagnosis, due to the lack of studies proving their pathogenicity.

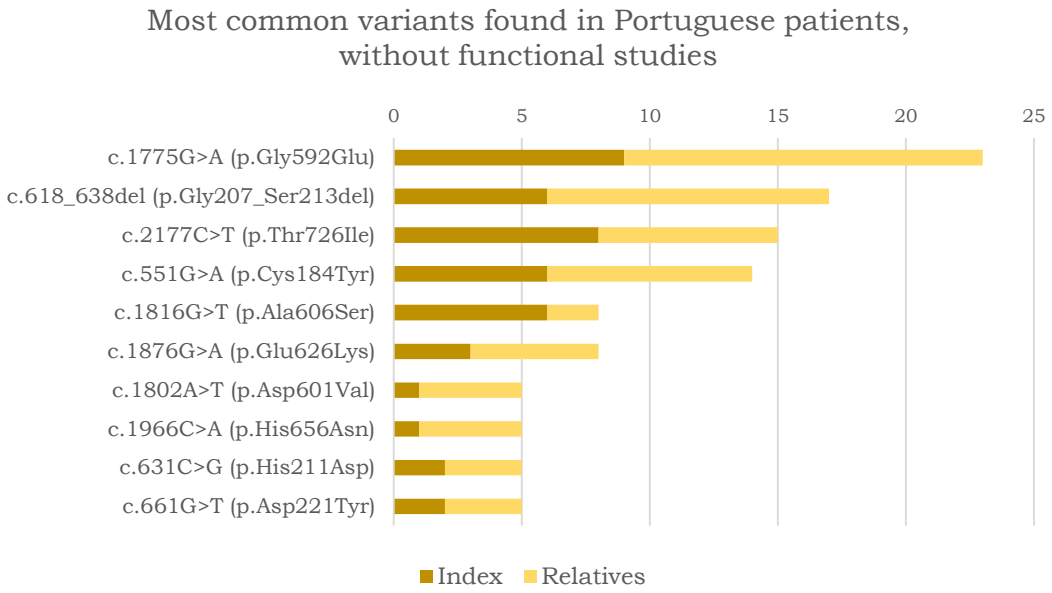


Figure 3.3. Number of patients carrying the most common variants identified in Portuguese FH Study, *without functional assays.* *Source: Portuguese Familial Hypercholesterolemia Study.*

Although none of the ten variants under study was found in a Portuguese normolipidemic panel constituted by 95 normolipidemic individuals (190 non FH alleles), it was possible to observe that variants c.2177C>T p.(Thr726Ile) and c.1966C>A p.(His656Asn) do not seem to co-segregate with the hypecholesterolemic phenotype in the families assessed (see *table 3.3.*). In addition, variant c.1816G>T p.(Glu626Lys) only co-segregates in 2 out of 3 families.

For all of them *in silico* prediction was performed (*table 3.3.*), classifying 7 of them as probably pathogenic, 1 as a neutral variant (c.2177C>T p.(Thr726Ile)) and 2 as variants of unknown significance (VUS) (c.1816G>T p.(Ala606Ser) and c.1816G>T p.(Glu626Lys)), as for these last, different software were in disagreement in their predictions. Furthermore, variants c.1775G>A p.(Gly592Glu), c.551G>A p.(Cys184Tyr), c.1966C>A p.(His656Asn), c.631C>G p.(His211Asp) and c.661G>T p.(Asp221Tyr) presented high PhyloP values, suggesting high conservation among different species.

Concerning structural location, and similarly to the variants found during molecular diagnosis, the majority of variants occurred either in the EGF precursor homology domain (6/10) or in the ligand-binding domain (4/10). The remain is placed in the O-linked sugars domain (c.2177C>T p.(Thr726Ile)).

Table 3.3. Characteristics and in silico prediction of the most common variants found in Portuguese population, without functional studies to date.

Alteration					First described by	Coseg	Portuguese Normolipidemic panel	*In silico* Prediction							Classification
Number of patients	Localization	cDNA	protein	Domain				Mutation Taster	Polyphen 2	SIFT	PhyloP	Human Splicing Finder	NNSSP	FSPLICE	
23	Exon 12	c.1775G>A	p.(Gly592Glu)	EGF precursor homology	[15]	17/18;0/3	0/190 non FH alleles	Disease causing	Possibly Damaging	Deleterious	5.811	100 (12A), 100 (12D), 90 (novel A)	100 (12A), 100 (12D)	100 (12A), 100 (12D)	Probably pathogenic
17	Exon 4	c.618_638del	p.(Gly207_Ser213del)	Ligand-binding	[72]	14/15;0/7	0/190 non FH alleles	Disease causing	NA	Deleterious	NA	100 (4A), 100 (4D)	100 (4A), 100 (4D)	100 (4A), 100 (4D)	Probably pathogenic
15	Exon 15	c.2177C>T	p.(Thr726Ile)	O-linked sugar	[15]	13/14;2/4	0/190 non FH alleles	Polymorphism	Benign	Tolerated	1.027	100 (15A), 100 (15D)	100 (15A), 100 (15D)	100 (15A), 100 (15D)	Neutral
14	Exon 4	c.551G>A	p.(Cys184Tyr)	Ligand-binding	[82]	8/10;0/0	0/190 non FH alleles	Disease causing	Probably Damaging	Deleterious	5.438	100 (4A), 100 (4D)	100 (4A), 100 (4D)	100 (4A), 100 (4D)	Probably pathogenic
8	Exon 12	c.1816G>T	p.(Ala606Ser)	EGF precursor homology	[83]	7/7;0/2	0/190 non FH alleles	Disease causing	Benign	Neutral	1.373	100 (12A), 100 (12D)	100 (12A), 100 (12D)	100 (12A), 100 (12D)	VUS
8	Exon 13	c.1876G>A	p.(Glu626Lys)	EGF precursor homology	[84]	7/10;1/3	0/190 non FH alleles	Disease causing	Benign	Deleterious	1,82	100 (13A), 100 (13D)	100 (13A), 100 (13D)	100 (13A), 100 (13D)	VUS
5	Exon 12	c.1802A>T	p.(Asp601Val)	EGF precursor homology	[58]	4/5;0/4	0/190 non FH alleles	Disease causing	Probably Damaging	Deleterious	2.104	100 (12A), 100 (12D), 82 (novel A), 102 (novel D)	100 (12A), 100 (12D)	100 (12A), 100 (12D)	Probably pathogenic
5	Exon13	c.1966C>A	p.(His656Asn)	EGF precursor homology	[85][86]	2/3;3/7	0/190 non FH alleles	Disease causing	Possibly Damaging	Deleterious	5.766	100 (13A), 100 (13D)	100 (13A), 100 (13D)	100 (13A), 100 (13D)	Probably pathogenic
5	Exon 4	c.631C>G	p.(His211Asp)	Ligand-binding	[72]	4/4;0/4	0/190 non FH alleles	Disease causing	Probably Damaging	Deleterious	5.864	100 (4A), 100 (4D)	100 (4A), 100 (4D)	100 (4A), 100 (4D)	Probably pathogenic
5	Exon 4	c.661G>T	p.(Asp221Tyr)	Ligand-binding	[86]	5/5;0/2	0/190 non FH alleles	Disease causing	Probably Damaging	Deleterious	5.864	100 (4A), 100 (4D)	100 (4A), 100 (4D)	100 (4A), 100 (4D)	Probably pathogenic

Coseg, co-segregation in families of the Portuguese Study of Familial Hypercholesterolemia, variant carriers/total affected; variant carriers/total non-affected; NA, not applicable; *In silico* prediction: SIFT, sort intolerant from tolerant; NNSSP, Nearest-neighbor Secondary Structure prediction; FSPLICE, find splice sites in genomic DNA A, Acceptor; D, Donor; VUS, Variant of Unknown Significance.

Two of these variants, c.2177C>T p.(Thr726Ile) and c.618_638del p.(Gly207_Ser213del), were found in patients during the molecular diagnosis performed in the first part of this project, so their family trees are shown in *figure 3.4*. However, no relatives' samples were, to date, available for the performance of cascade screening.

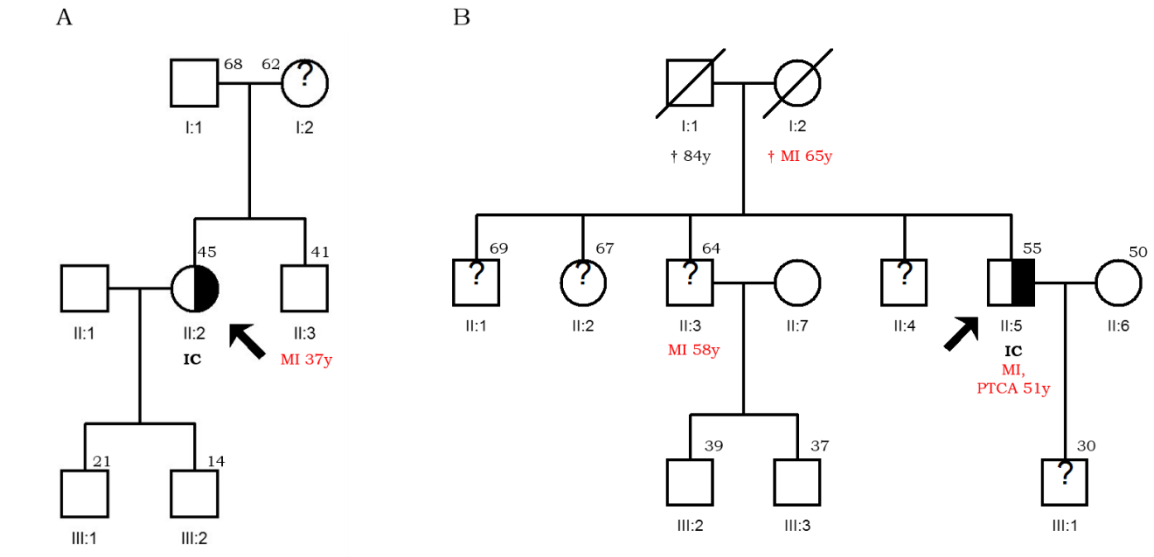


Figure 3.4. Family trees of index cases (IC) carrying variants (A) c.2177C>T p.(Thr726Ile) and (B) c.618_638del p.(Gly207_Ser213del). *Arrows indicate index cases. Question marks indicate reported hypercholesterolemic relatives, in whom molecular diagnosis was not performed, due to lack of DNA samples. MI, Myocardial infarction; PTCA, Percutaneous transluminal coronary angioplasty.*

It is possible to observe that, in family A, only two people are hypercholesterolemic. Although II:3 have suffered a MI at young age but, to our knowledge, his cholesterol values are normal. Furthermore, patient II:2, in who this variant was identified, have not suffered from heart disease, as well as her mother, suggesting that this variant may not be the cause of their hypercholesterolemic phenotype.

In family B is possible to observe a high prevalence of the hypercholesterolemic phenotype. Furthermore, two of the alive hypercholesterolemia patients have suffered MI, suggesting that this variant may be the cause of their aggressive hypercholesterolemic phenotype.

Results from functional studies of both variants are needed for the correct assessment of these variants' pathogenicity.

Site-directed mutagenesis

Each of the nucleotide changes, responsible for the variants c.1775G>A p.(Gly592Glu), c.618_638del p.(Gly207_Ser213del), c.2177C>T p.(Thr726Ile), c.551G>A

p.(Cys184Tyr), c.1816G>T p.(Ala606Ser), c.1876G>A p.(Glu626Lys), c.1802A>T p.(Asp601Val), c.1966C>A p.(His656Asn), c.631C>G p.(His211Asp), c.661G>T p.(Asp221Tyr) and c.1633G>T p.(Gly545Trp) (control), were successfully introduced by site-directed mutagenesis in the pnDNA3 plasmid containing the LDLR cDNA.

Most of the times, no bands were observed after the mutagenesis PCR. However, as a non-observable small amount of plasmid DNA is enough to transform bacteria, the experiment proceeded. Thus, the presence of desired variants and the integrity of the insert of the *LDLR* gene, was successfully confirmed by Sanger sequencing. Even though recloning was well succeeded only in 9/11 variants (*figure 3.5.*), all of them proceeded for functional assessment. This decision will be further addressed in section *4.2. Functional Studies.*

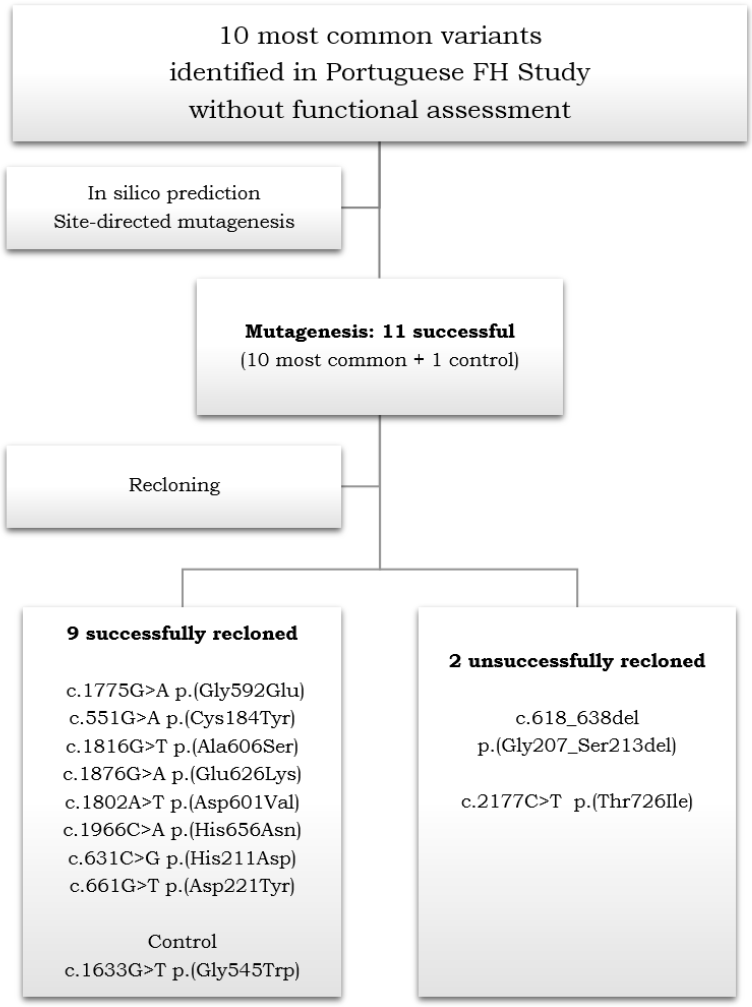


Figure 3.5. Schematic representation of site-directed mutagenesis and recloning results. *While site-directed mutagenesis was successfully performed for all variants, recloning was not successful in 2 of these variants.*

3.3. Functional studies

3.3.1. Lipoprotein labelling with FITC

After ultracentrifugation of human plasma samples, several bands corresponding to distinct lipoproteins were observable. This occurred due to the density gradient created with KBr, as described in Methods. As result, after ultracentrifugation, all particles in solution were distributed vertically in the tube according to their density (*figure 3.6.*).

Three bands were observable: the first one, at the top of the tube, were VLDL, as these are the less dense lipoproteins. A bit lower than the middle of the tube, LDL band can be seen, which was the fraction isolated to proceed for fluorescent labelling. Almost at the bottom of the tube, it was possible to observe a band corresponding to HDL, as these are the densest lipoproteins.

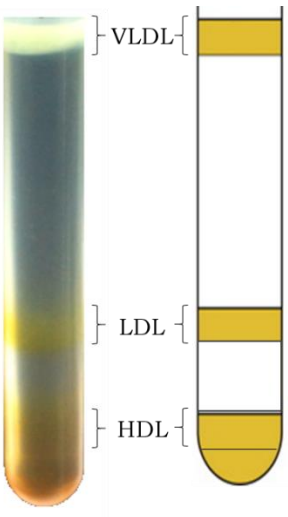


Figure 3.6. Major human plasma lipoprotein fractions after one-step salt gradient ultracentrifugation with KBr. *Three bands, corresponding to VLDL, LDL and HDL fractions, were observed. (Adapted from Alves A.C., 2014, "Base genética da Hipercolesterolemia Familiar", Universidade de Lisboa)*

On average, 1 mg of LDL was obtained, after purified and assayed for protein content. Labelling with FITC occurred successfully and samples were correctly visualised in the flow cytometer.

3.3.2. FACS assays

During the flow cytometric assays, a well-defined population of CHO-ldlA7 cells was observed (*figure 3.7.A and B.*). Fluorescence associated to the LDLR expression at the cell surface, and to the LDLR binding and uptake activities was measured in these cells (*figure 3.7.C.*).

The population under study is selected into the blue lines. Events observed outside these lines correspond to cells which were lysed during the transfection process or the sample scraping. The integrity of lysed cells is compromised, thus correspondent SSC and FSC values are lower (*figures 3.7.A.* and *B.*) and these cells were not taken into account for the analysis.

In *figure 3.7.C.* is possible to observe the number of cells in relation to fluorescence measures. Here, two types of peaks are observable: one dislocated to the left (at lower fluorescence values), and the second one dislocated to the right (at higher fluorescence values).

The first peak, dislocated to the left, at lower values of fluorescence, corresponds to the basal fluorescence of cells. A population of untransfected cells, represented by the yellow line (*figure 3.7.C*), only presents this peak, as no LDLR is detected and only basal fluorescence is measured. The population of cells transfected with the plasmid carrying the variant p.(Gly545Trp), represented by the red line (*figure 3.7.C*), which results in no expression of LDLR at the cell membrane surface, also presents a unique peak, corresponding to the basal fluorescence. A second peak, dislocated to the right, is observable in a cell population, which presents additional fluorescence, as a population transfected with a wt plasmid (green line, *figure 3.7.C*). This additional fluorescence occurs due to the LDLR at the cell surface, immunodetected with antibodies bound to Alexa-488 fluorophore (for expression). For binding and uptake assays, the same kind of graphic representation was obtained, but here the fluorophore used was FITC.

Data analysis for all variants and respective graphic representations (*figure 3.8.*) were performed as described in Chapter 2. Materials and Methods, section *2.3.2.4. Measurements by Flow Cytometry*.

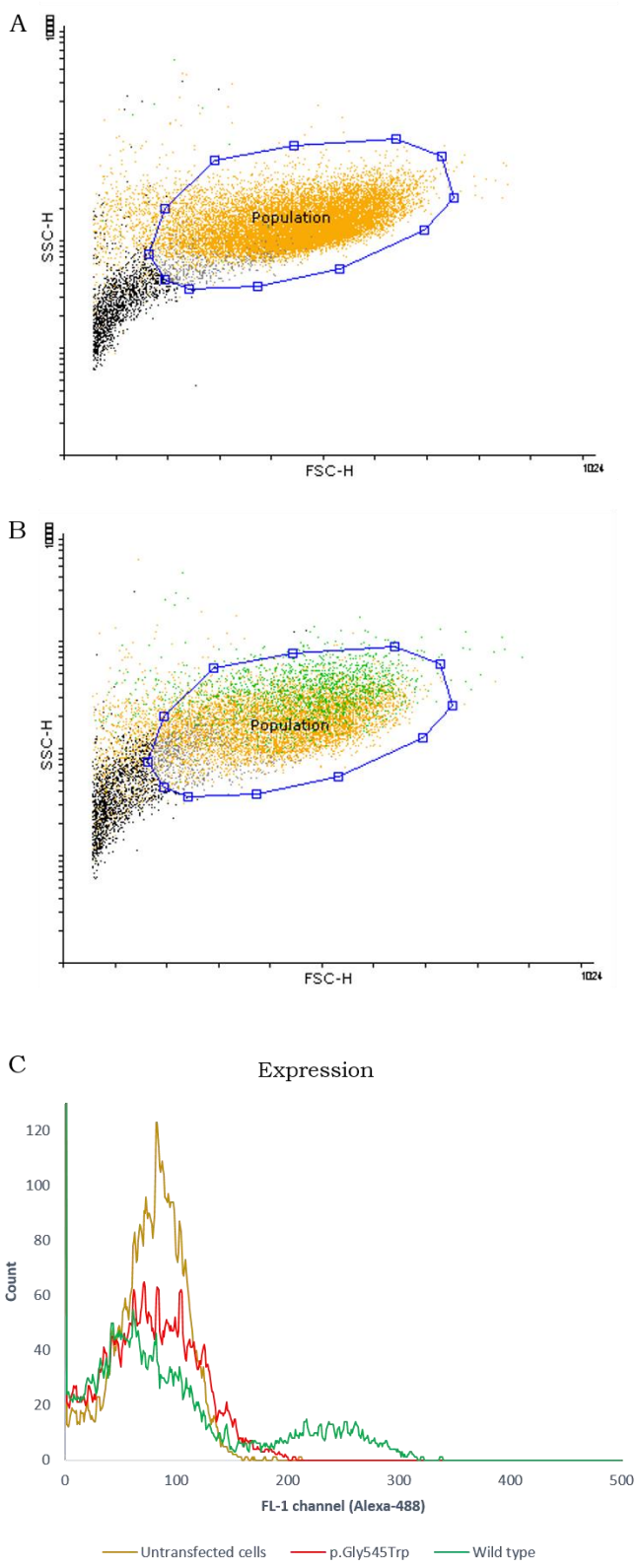


Figure 3.7. Flow cytometric outputs – example of LDLR expression in CHO-ldlA7 cells. *A well-defined cell population is observed in (A) and (B). In (B), cells transfected with the wt plasmid and expressing LDLR at membrane surface (green dots) present higher fluorescence values, due to Alexa Fluor 488-conjugated goat anti-mouse IgG. These fluorescence values are represented by the green line in (C). In addition to basal fluorescence (yellow and red lines), these cells presented fluorescence due to the LDLR expression, presenting a second peak (green line), dislocated to the right.*

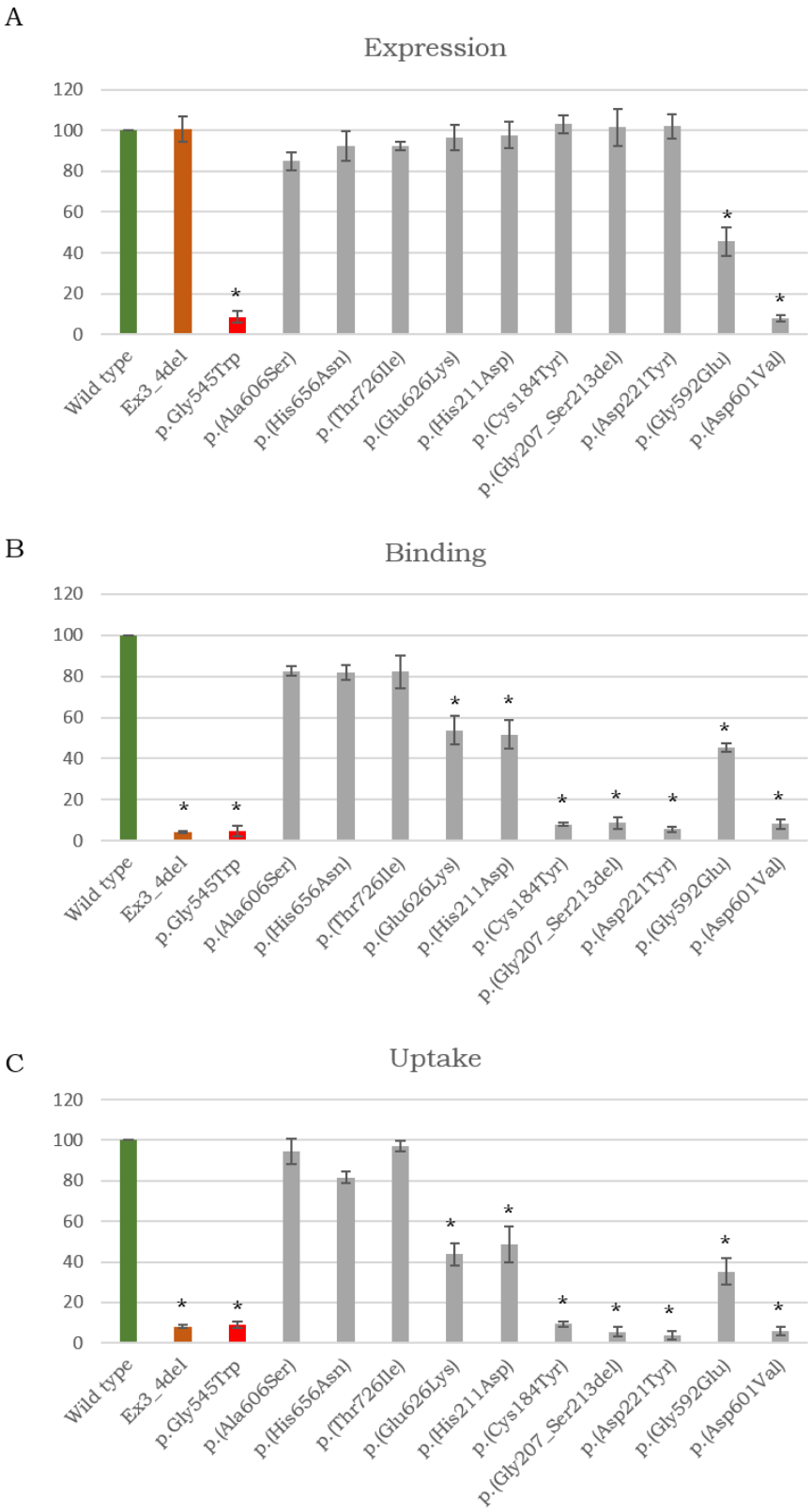


Figure 3.8. Functional characterization of LDLR variants in transfected CHO-ldlA7 cells.
*(A) LDLR expression at the cell membrane. (B) FITC-LDL binding after 4h incubation at 4ºC. (C) FITC-LDL uptake after 4h incubation at 37ºC. Bars represent values as the mean of three independent measurements. Error bars represent +/- standard deviation (SD). *p-value<0.01 compared to LDLR WT using Student's t-test.*

In *figure 3.8.* is represented the final results for cell surface LDLR expression, LDL-LDLR binding and uptake. Three transfected cell controls were used: with the wt plasmid, with Ex3_4del plasmid, which expresses but does not binds or uptake LDL, and with p.(Gly545Trp) plasmid, which does not express and consequently does not bind or uptake LDL. All values were normalized for wt plasmid transfection, which was considered 100%.

Cells transfected with plasmids carrying variants p.(Ala606Ser), p.(His656Asn) and p.(Thr726Ile) presented cell surface expression, binding and uptake values comparable to cells transfected with the wt plasmid. Thus, they were considered as neutral variants.

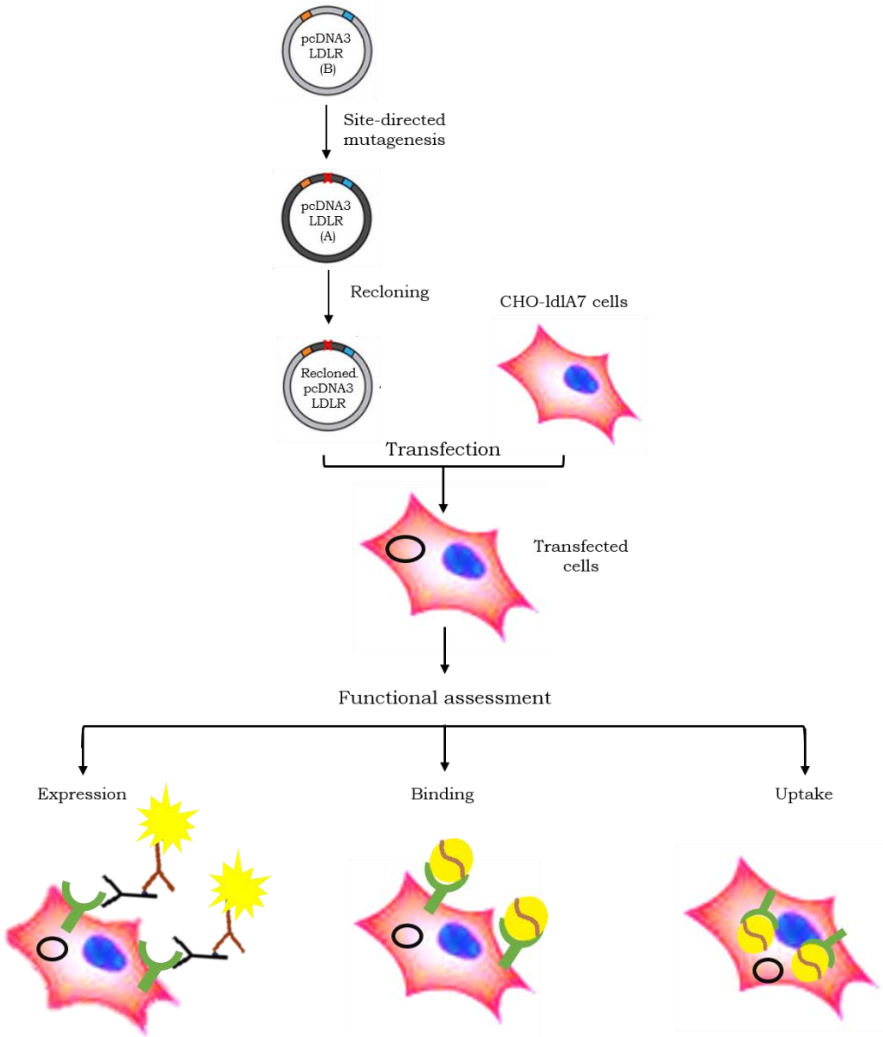
Cells transfected with plasmids carrying variants p.(Glu626Lys) and p.(His211Asp) presented normal expression at the cell surface. However, only approximately 50% of binding activity was observed. Consequently, uptake values also decrease for approximately 50%. Thus, the presence of these variants seem to lead to a partial impairment of the LDLR binding activity, being considered pathogenic.

Variants p.(Cys184Tyr), p.(Gly207_Ser213del) and p.(Asp221Tyr) also seem to cause an impairment of LDLR binding activity, but in a more aggressive way. Although the LDLR seems to be normally expressed, these variants presented very low binding and uptake activities, comparable to the control Ex3_4del. Therefore, variants p.(Cys184Tyr), p.(Gly207_Ser213del) and p.(Asp221Tyr) lead to an impairment of LDLR ability of LDL binding and uptake, also being considered pathogenic.

Variant p.(Gly592Glu) presented a milder value, of approximately 50%, for cell surface expression. Consequently, binding and uptake activities also presented approximately half of the activities observed in cells transfected with the wt plasmid. These results lead to the assumption that, although not totally, LDLR function is affected by the presence of this variant. Nevertheless, this variant was considered pathogenic.

Variant p.(Asp601Val) does not express LDLR at the cell surface, so binding and uptake activities are also compromised. This results in an almost total impairment of the receptor activity, being considered pathogenic.

Concluding, among ten variants under study, three (variants p.(Ala606Ser), p.(His656Asn) and p.(Thr726Ile)) did not reveal any impact in the LDLR function. However, the remaining 7 variants assessed seem to affect the LDLR activity. Final values, in percentage, are shown in *figure 3.9.*, along with a schematic representation of results for the three parameters here assessed (expression at the cell surface, and LDL binding and uptake). It is important to refer that values between 80% and 100% were considered to be associated to a complete activity of the receptor, while values lower than 80% were associated to a compromised activity of the receptor. Furthermore, values lower than 2% were considered to be associated with null variants.



Variant	Expression	Binding	Uptake
Wild type (control)	100%	100%	100%
Ex3_4del (control)	80-100%	<2%	<2%
p.(Gly545Trp) (control)	<2%	<2%	<2%
c.1816G>T p.(Ala606Ser)	79-91%	79-85%	86-100%
c.1966C>A p.(His656Asn)	82-96%	76-85%	79-86%
c.2177C>T p.(Thr726Ile)	89-95%	71-90%	94-100%
c.1876G>A p.(Glu626Lys)	90-104%	45-62%	37-50%
c.631C>G p.(His211Asp)	92-106%	43-60%	36-55%
c.551G>A p.(Cys184Tyr)	96-106%	<2%	<2%
c.618_638del p.(Gly207_Ser213del)	92-110%	<2%	<2%
c.661G>T p.(Asp221Tyr)	96-109%	<2%	<2%
c.1775G>A p.(Gly592Glu)	39-53%	43-48%	30-45%
c.1802A>T p.(Asp601Val)	<2%	<2%	<2%

Figure 3.9. *Schematic representation of characteristics assessed by flow cytometry and values obtained for each variant under study. Expression is assessed through the use of the fluorophore Alexa-488 linked to antibodies immunodetecting LDLR at the membrane surface; binding and uptake are assessed using FITC-labelled LDL, measured at the surface and in the interior of cell, respectively. Values are resultant of three independent experiments, in triplicate.*

3.4. *In silico* vs. *in vitro*

According to results observed, it can be inferred that all variants, except p.(Ala606Ser), p.(His656Asn) and p.(Thr726Ile) seem to lead to an impaired LDLR function.

Comparing functional results for all variants of the present work with *in silico* prediction (*table 3.4.*), it is possible to observe that all variants were correctly predicted, except one. Variant c.1966C>A p.(His656Asn), proved to be neutral, was predicted to be probably pathogenic.

Table 3.4. *Comparison between in silico prediction and functional assessment previously reported, for variants identified in molecular diagnosis, and here reported, for variants assessed during this project.*

Variant				***In silico* prediction**	**Functional assessment**	**References**
Localization	**cDNA**	**protein**	**Domain**			
APOB Exon 26	c.10580G>A	p.(Arg3527Gln)	Domain 6	Probably pathogenic	Pathogenic	[76]
LDLR Exon 3	c.301G>A	p.(Glu101Lys)	Ligand-binding	Probably pathogenic	Pathogenic	[15]
LDLR Intron 3	c.313+1G>A	p.Leu64_Pro105delinsSer /p.Pro105_Ala860delinsArg LysCysGlyProAlaPheAlaIleG luProIle	EGF precursor homology	Probably pathogenic	Pathogenic	[78][79][80]
LDLR Exon 4	c.589T>C	p.(Cys197)Arg	Ligand-binding	Probably pathogenic	Pathogenic	[77]
LDLR Exon 4	c.670G>A	p.(Asp224Asn)	Ligand-binding	Probably pathogenic	Pathogenic	[15]
LDLR Intron 7	c.1060+1G>A	skipping of exon 7 (p.Gly314Aspfs*507)	EGF precursor homology	Probably pathogenic	Pathogenic	[74]
LDLR Exon 11	c.1633G>T	p.(Gly545Trp)	EGF precursor homology	Probably pathogenic	Pathogenic	[73]
LDLR Exon 4	c.551G>A	p.(Cys184Tyr)	Ligand-binding	Probably pathogenic	Pathogenic	During the present work
LDLR Exon 4	c.631C>G	p.(His211Asp)	Ligand-binding	Probably pathogenic	Pathogenic	During the present work
LDLR Exon 4	c.618_638del	p.(Gly207_Ser213del)	Ligand-binding	Probably pathogenic	Pathogenic	During the present work
LDLR Exon 4	c.661G>T	p.(Asp221Tyr)	Ligand-binding	Probably pathogenic	Pathogenic	During the present work
LDLR Exon 12	c.1775G>A	p.(Gly592Glu)	EGF precursor homology	Probably pathogenic	Pathogenic	During the present work
LDLR Exon 12	c.1802A>T	p.(Asp601Val)	EGF precursor homology	Probably pathogenic	Pathogenic	During the present work
LDLR Exon 12	c.1816G>T	p.(Ala606Ser)	EGF precursor homology	VUS	Neutral	During the present work
LDLR Exon 13	c.1876G>A	p.(Glu626Lys)	EGF precursor homology	VUS	Pathogenic	During the present work
LDLR Exon13	c.1966C>A	p.(His656Asn)	EGF precursor homology	Probably pathogenic	Neutral	During the present work
LDLR Exon 15	c.2177C>T	p.(Thr726Ile)	O-linked sugar	Neutral	Neutral	During the present work

For variants c.1816G>T p.(Ala606Ser) and c.1816G>T p.(Glu626Lys), predictions of different *in silico* tools were not in agreement. For this reason these variants were classified as VUS, not being possible to compare this prediction to functional studies results.

In a total of 14/17 variants, the *in silico* prediction was able to correctly identify the pathogenicity, leading to 82% correct identification rate. The results suggest that *in silico* tools contribute for variant classification when no functional studies are available.

Chapter 4

Discussion

4.1. Molecular diagnosis

Looking at the biochemical parameters of the 25 patients, some dispersion of values is observable, not being possible to distinguish individually between patients with and without a variant only based in their lipid profile. This leads to the assumption that FH patients can only be correctly identified when molecular diagnosis is performed, as their lipid profile is not enough to predict the existence of a genetic defect. In adults, molecular diagnosis becomes even more important, since environmental factors can interfere even more with biochemical parameters.

Only 8/25 patients were identified with a disease-causing variant - 2 patients were identified with a variant in *APOB* and 6 with variants in *LDLR*. Furthermore, four variants with no functional studies proving their pathogenicity were also found and, if *in silico* predictions were correct, 2/4 would be disease-causing. This would lead to 10/25 patients identified with a disease-causing variant (40%), a value that would be in agreement with literature (41.5%) [58].

In adult group the positive rate was extremely lower that the one observed in the pediatric group, in which a positive rate of 78% was observed. This can be explained by environmental factors, as at adult age, several factors can influence cholesterol values, leading to a false FH clinical diagnosis [87]. In children, it can be claimed that environmental factors could not have had the time to influence cholesterol levels. This justification implies that children present a true phenotype, not influenced by environmental factors, which can constitute a more faithful representation of FH genotype.

The distinction between a monogenic dyslipidemia and an environmental dyslipidemia plays an important role in cardiovascular risk assessment and treatment of

these patients [87]. While a monogenic condition is associated with a higher cardiovascular risk, polygenic environmental dyslipidemia is associated with a risk that may be reduced only with the implementation of lifestyle measures and counselling [88][89].

Patients identified with variants described as pathogenic

In this cohort, 7 variants already proved as disease-causing, were found in 8 index cases (patients 2 - 6, 8, 9, 15), providing them a definitive diagnosis.

Variants c.10580G>A p.(Arg3527Gln) in *APOB* and c.301G>A p.(Glu101Lys), c.589T>C p.(Cys197Arg), c.670G>A p.(Asp224Asn), and c.313+1G>A in *LDLR* had already been described in other European countries [15], [76], [78], [90]. Nevertheless, variants c.1633G>T and c.1060+1 G>A have only been described in the Portuguese population [72] to date. Furthermore, these last present functional studies proving their pathogenecity [73], [74], as a result of the effort to fully characterize all variants found in Portugal [58].

For patients identified with variants previously described as pathogenic, it is possible to give a definite diagnosis of FH, as variants found can be considered the justification for their phenotype.

A definitive diagnosis allows early personalized counselling and treatment, improving these patients' prognosis. With appropriate dietary and lifestyle advice and adequate therapeutic measures, it will be possible to reduce the risk of premature CHD.

Comparing *in silico* predictions with the results obtained from functional studies, it is possible to observe that all variants proved as disease-causing were predicted to be pathogenic by prediction tools here used. Nevertheless, these prediction tools should be used carefully. Their advantages and limitations will be further discussed in section *4.4. In vitro* vs. *In silico*.

Patients identified with variants of unknown significance

Four variants (c.618_638del p.(Gly207_Ser213del), c.2177C>T p.(Thr726Ile), c.1585G>C p.(Gly529Arg) and c.1186+56_1186+64del), which role in LDLR function is unknown, as they have no functional studies proving their pathogenicity, were found. For these patients a definitive diagnosis of FH was not possible, corroborating the importance of performing *in vitro* studies for reported variants, which effect on FH phenotype is still unknown.

Functional assessment of these variants, in order to prove their influence in the LDLR activity, becomes imperative, to provide a definite diagnosis for FH patients.

Variants c.618_638del p.(Gly207_Ser213del) and c.2177C>T p.(Thr726Ile) constitute two of the most common missense variants in Portuguese population without functional characterization. For this reason, they are included in the group of variants functionally assessed in this work (see section *3.2. Production of LDLR variants*) and results regarding their pathogenicity will be further discussed in section *4.3. Functional Studies.*

When functional assessment is performed for variants with unknown pathogenicity, two situations can be faced: if these variants were proved as pathogenic, patients would have a definitive diagnosis of FH, as its pathogenicity would be the cause of their phenotype; nevertheless, these variants might as well be proven as neutral, not constituting the cause of patients' FH phenotype.

Patients carrying neutral variants might have other justification for their phenotype, similarly to patients in whom no variant was found.

Patients with no variant found

Among the group of patients under study during the present work, 2 children and 11 adults remain without any identified variant associated to their phenotype.

It is worth mentioning that only fragments of exons 26 and 29 of *APOB* were studied and pathogenic variants outside these fragments have been previously reported [41] as the cause of FH. Moreover, *PCSK9* was not studied during this project, and variants in this gene can be the cause of FH phenotype, as previously reported [38]. Thus, the study of complete *PCSK9* and *APOB* genes, may harbour evidence supporting a definitive FH diagnosis for these patients.

Although variants in *LDLR*, *APOB* and *PCSK9* genes are, currently, the genes associated to FH, variants identified in other genes may be the cause of a hypercholesterolemic phenotype. For instance, a deletion in *APOE*, which is a lipoprotein to which LDLR has high affinity, has been associated with FH [91]. Furthermore, variants of the Niemann-Pick C1-Like 1 gene (*NPC1L1*), which encodes a protein implied in intestinal sterol absorption, have also been associated with FH [92].

The cause of these patients' phenotype may as well rely in other dyslipidaemias, such as familial combined hyperlipidemia (FCHL). FCHL is a polygenic disease, for which a clinical diagnosis is not possible, as the phenotype can vary among family members and over time [93]. Although some individuals could also present a clinical criteria for FCHL (patients 10, 14, 15, 17, 19, 22 and 25), FH should always be ruled out first, since a genetic diagnosis of FCHL does not exist or it is not possible. [93]

The remaining 2 cases (patients 1 and 7) are the only two children in who no variant was identified. However, in these patients, an ApoB/ApoAI ratio below 0.68 was observed. This is in agreement with literature, as a ratio above 0.68 has been proved to be a clinical criteria for FH in children [87]. These results suggest that, in addition to Simon Broome criteria, the ApoB/ApoAI ratio seems to be an effective criteria in the identification of hypercholesterolemic children.

Furthermore, patient 1 has high Lp(a) values, a small lipoprotein structurally similar to LDLR [9][10], which is independently associated with CHD risk [94]. Recently evidence relating its catabolism with PCSK9 and LDLR have also been provided [95], which indicates the possibility of this lipoprotein playing a role in cholesterol levels.

According to the international FH foundation, the most reliable diagnosis of FH can be made using both phenotypic and genetic testing [59]. Results here obtained suggest that FH patients may be clinically misdiagnosed due to environmental factors, mainly in adults, and DNA testing increases the accuracy of detecting FH [59].

Furthermore, DNA testing makes cascade screening more cost effective [96], and should be used to screen family members after the variant is identified in index cases [59]. Although only 9 individuals (related to six index cases) were studied during this project, cascade screening should be performed as soon as DNA samples from remain relatives are available. This will allow the correct counselling and implementation of lifestyle measures to relatives who carry disease-causing variants.

If genetic testing detects a variant, its significance as a disease-causing variant needs to be assessed and recorded [59], namely by the performance of functional studies. For this reason, in the scope of the Portuguese FH Study, the most common alterations found in this cohort, without functional studies proving their pathogenicity are currently under study.

4.2. Functional Studies

The ten most common variants found in Portuguese patients were created in a pcDNA3_LDLR plasmid, in order to assess their impact in LDLR function. Among these, here functionally assessed, 2 (c.618_638del p.(Gly207_Ser213del) and c.2177C>T p.(Thr726Ile)) were identified in patients during the first phase of this project, in molecular diagnosis. Moreover, the variant here used as negative control, (c.1633G>T p.(Gly545Trp)), was also identified in one of the studied individuals. These results show that, in addition to novel variants, previously reported variants are often identified in patients.

Genetic bases of FH in the Portuguese population are heterogeneous [57][58][72][74], as well as in other studied populations [84][85][97], having been found several types of variants spread throughout the *LDLR* gene.

Although more than 1695 *LDLR* variants have been identified [98], the majority of them do not have functional studies proving their pathogenicity. Thus, functional assessment of these variants becomes imperative, as part of the genetic FH diagnosis. During the past years, several methods have been used, all with the same purpose: functionally assess of *LDLR* variants.

Under the usual conditions of cell culture, human and animal cell do not synthesize their own cholesterol, but rather use the LDLR pathway to derive cholesterol from LDL.

Therefore, the LDL pathway was elucidated entirely through studies of cultured human skin fibroblasts [99].

Receptor assays reported so far include the use of radiolabeled-LDL or fluorescently-labeled LDL to measure the LDLR activity in skin fibroblasts [99][100], leukocytes [101][102], and heterologous cells [69]. Furthermore, a strain of rabbits, designated Watanabe heritable hyperlipidemic rabbits (WHHL rabbits) represents an animal model for the study of FH, as they present a homozygous pathogenic variant in the *LDLR* and also develop coronary and aortic atherosclerosis [103]. Nevertheless, work with animal models has complications and costs associated. However, a sample of skin or blood, for fibroblasts or lymphocyte culture, respectively, is not always available at the time of the functional characterization. For this reason, heterologous cells transformed with *in vitro* mutated plasmids constitute the chosen method for functional assessment in this project.

Regarding the labelling, the reference method to estimate the LDLR activity was, for many years, the radioactive assay with 125-Iodine (125I) [104]. Still, approaches using fluorescently-labelled LDL have later been described, associated with flow cytometry [73][105][106][107]. Comparing both labelling methods [69], although the radioactive assay is very sensitive, the use of fluorimetric assays based on covalent labelling of LDL with fluorophores such as FITC overcomes many of the problems associated with radiolabelling (such as the risk of exposure of researchers to radioisotopes or the difficulties and ethical considerations of nuclear waste elimination procedures). Moreover, FITC-labelling LDL constitutes a much cheaper procedure.

Thus, fluorescence-based methodology constitutes an accurate, cleaner and cheaper methodology for *LDLR* variants functional assessment, mainly in times where the development of new techniques as next generation sequencing (NGS) provides a high number of information and identified variants, some of them still in need of functional validation.

All site-director mutagenesis occurred successfully. However, although recloning only was successful in 8/10 variants under study, all ten variants proceeded for functional characterization. This decision was based in the fact that the enzyme used for the site-directed mutagenesis was a PfuUltra High Fidelity DNA polymerase, which presents an error rate of 1.3 to 2.8×10^{-6}, 10x lower than the error rate observed with Taq polymerase [108][109]. Actually, pfu polymerases' error rate is so low that could raise questions about the necessity of the recloning step, which is time consuming and with potential technical difficulties [110]. However, as it is an additional step that assures the integrity of the plasmid, it was performed whenever possible.

Functional characterization for the determination of pathogenicity was based in two types of essay: the measurement, in specific conditions, of FITC-labelled LDL binding and uptake; and the determination of mature LDLR levels at cellular surface through

immunocytochemistry, with specific antibodies for the receptor. Both were performed in CHO-A7 cells, transfected with different LDLR variants under study.

Information obtained with these assays allows inferences about the relation between the variant and the protein activity, leading to a prediction of the severity of phenotype associated to this variant.

Furthermore, an analysis of the currently existing databases [111][112][113], for *LDLR* variants, was performed, in order to compare and better discuss results here obtained.

4.2.1. Neutral Variants

Variants c.1816G>T p.(Ala606Ser), c.1966C>A p.(His656Asn) and c.2177C>T p.(Thr726Ile) were classified as neutral variants, as observed activities were similar to those observed in wt.

Variant c.1816G>T p.(Ala606Ser) is located in the EGF precursor homology domain. Although this is a highly conserved domain among species, this variant presents a low phyloP, revealing a poor conservation. Thus, it is expectable that a poorly conserved position may not play an important role at the level of protein function.

Nevertheless, variant c.1966C>A p.(His656Asn) presents a high phyloP value, indicating conservation among species, which should not be expected for a neutral variant. However, this variant seems to not co-segregate with the phenotype in families studied, which corroborates its neutrality regarding FH.

Variant c.2177C>T p.(Thr726Ile), located in the O-linked sugars domain, also has a low phyloP value and seems to not co-segregate in studied families. Previously, some studies regarding this variant have been performed, indicating its present in normlipidemic individuals [114]. However, functional studies were only performed with a compound heterozygous patient's lymphocytes and, although it had resulted in a LDLR activity of 2-5% [15], it was not clear which variant was responsible for these results. Besides that, results in here show its neutrality and it was also predicted to be neutral by *in silico* prediction tools.

Furthermore, variant c.2177C>T p.(Thr726Ile) was found in patient 23, during molecular diagnosis. This patient's TC and LDL-C values before medication were border line (see section *3.1. Molecular diagnosis*), when taking into account the criteria used for clinical diagnosis. Although this patient presented family history of high cholesterol, their border line values may be easily justified by environmental factors.

Most possibly, the changes caused by these three variants are not enough to create a conformational alteration in the protein, or at least an aggressive one, sufficient to compromise the receptor function. Thus, it is expectable that patients carrying these variants may have other justification for their phenotype (see section *4.1. Molecular*

Diagnosis), as they do not seem to be pathogenic and *LDLR* seem to be correctly expressed and positioned at the membrane surface and efficiently bind and uptake LDL.

4.2.2. Null variants

Variant c.1802A>T p.(Asp601Val), located at the EGF precursor homology domain, presented fluorescence values comparable to negative control. This variant has only been reported in the Portuguese population [58]. It is located in the EGF precursor homology domain, which is a domain extremely conserved among species, although it presented a medium phyloP value.

It is observable that the activity of the LDLR produced by this variant is comparable to the control, negative for LDLR expression at the cell surface, which leads to the assumption that the LDLR does not reach the cell membrane. However, with the method used, it was not possible to distinguish if LDLR was not synthesized at all, or if the variant was partially or totally retained in the ER. For this reason, the distinction between classes I, II A and II B, respectively, was not possible. These variants must be further studied through protein expression assessment by Western Blot, which would allow the distinction between the precursor and mature form of the receptor, as well as respective amounts. Confocal laser scan microscopy would also be a good technique to confirm all results obtained.

The correct protein synthesis and the lack of anchoring to the membrane could also be pointed, but it seems unlikely, as the variant is not present in the domains implied in membrane anchoring.

As this variant seems to lead to a complete lack of LDLR function, inclusively compared to the negative control for the *LDLR* expression, it seems correct to affirm that this variant is pathogenic. Thus, in a heterozygous patient carrying this variant, only approximately 50% of LDLR will be functionally normal, being expected a severe FH phenotype. In the case of an individual be homozygous for this variant, no receptors would be synthesized and this would result in a brutally increased cholesterol levels and so an increased cardiovascular risk at early ages.

4.2.3. Binding-defective variants

Variants c.661G>T p.(Asp221Tyr), c.618_638del p.(Gly207_Ser213del) and c.551G>A p.(Cys184Tyr), presented fluorescence values for binding and uptake comparable to negative controls for binding and uptake activities. All of them are located at the LDL-binding domain, where variants often tend to impair LDLR function [15], due to the incapability of the mutant LDLR in recognizing the LDL.

Variant c.551G>A p.(Cys184Tyr) leads to the substitution of a cysteine, an amino acid present all over the repeats of ligand-binding domain, which play an important role in the structure of this domain [51].

Variant c.661G>T p.(Asp221Tyr) results of a change of an amino acid with a side chain negatively charged for a neutral amino acid. This may be enough to break interactions at the protein level, leading to changes in the recognition of LDL by the LDLR.

Variant c.618_638del p.(Gly207_Ser213del) was identified in patient 18 during molecular diagnosis, who presented family history of high cholesterol (see *table 3.1.* in Results, section *3.1. Molecular diagnosis*). Although it was only possible to register biochemical values on medication, the fact that he suffered a MI and was subjected to a percutaneous transluminal coronary angioplasty (PTCA) is enough to infer that this patient presents an aggressive phenotype. Furthermore, if a missense variant is capable of impairing the LDLR binding function, it is expectable that a deletion, even as if small as the one here studied, have a great impact at the protein function level.

As the occurrence of these three variants results in a complete lack of LDLR binding to the LDL, these cholesterol carriers are not withdrawn from the blood circulation. This results in increased cholesterol levels, probably leading to an aggressive phenotype, due to a severely reduced LDLR activity.

Variants c.631C>G p.(His211Asp) and c.1876G>A p.(Glu626Lys) presented milder fluorescence values for binding and uptake. To our knowledge, justifications laying in protein structure, for milder binding and uptake activities were not reported. However, some speculations can be made.

Variant c.631C>G p.(His211Asp) results in the change of a positively charged for a negatively charged amino acid and is located in the ligand-binding domain. This amino acid change can lead to the disruption of interactions in this domain. If the ligand-binding domain was somehow less stable, it could lead to the loss of part of bound LDL. This may be possible, because LDLR remains at the cellular surface for about 10 min [5] and, meanwhile it is not internalized, LDL could be freed due to this instability. However, the LDLR-LDL complex formation and rapid internalization could lead to the mild values here observed.

Variant c.1876G>A p.(Glu626Lys) is located in EGF precursor homology domain. As it leads to a change of a glutamate, negatively charged, for a lysine, which is positively charged, it is possible that some interactions are disrupted. It can be speculated that these disrupted interactions interfere with ligand-binding domain's stability, as when no LDL is present these two domains are considerably near to each other. In this case, similarly to variant c.631C>G p.(His211Asp), this instability could lead to a decreased amount of bound LDL.

Independently of all speculations, as these two variants, above mentioned, do not lead to a total impairment of these functions, they are expectable to cause milder phenotype in patients who carry them. It is important to refer that, if one of these variants is presented

in heterozygous and as it presents approximately 50% activity, in practice, the patient will only have a reduction of approximately 25% of the LDLR activity, as the healthy allele should be producing a normal and functional receptor and the FH allele only presents an impairment of approximately 50%.

As a result of an impairment of binding activity in these variants of LDLR, the uptake was also coherently affected in all cases. However, as uptake activities seem to be similar to binding activities into respective variants, it can be pointed that these variants do not cause any further effect at the uptake level, so lower uptake activities occur as a consequence of lower binding activities firstly performed.

4.2.4. Recycling-defective variants

The variant c.1775G>A p.(Gly592Glu), presented fluorescence values approximately of 50% when comparing to the wt expression, binding and uptake values. This variant occurred in exon 12, leading to an alteration at the protein level in the EGF precursor homology, more specifically in the β-propeller.

It has been proved that, although the ligand-binding domain is sufficient for binding lipoprotein particles, the EGF precursor homology domain is required for ligand release [18], [115], due to alterations in its conformation at acid pH in the endosome. Thus, a change in the β-propeller amino acid constitution can lead to an impairment of the LDL releasing, not being possible for the LDLR to be recycled to the membrane surface.

The LDLR remains at the cellular surface for about 10 min [5], being after internalized, even if no LDL is bound. Thus, while cells are incubated for 24 h in order to reach the maximal expression of the *LDLR*, cycle initiates and some LDLR are internalized. As recycling function is impaired in cells transfected with plasmids carrying this variant, only newly synthesized LDLR is present in the membrane and is immunodetected, leading to values approximately of half of the ones observed in wt.

So, it is expected that patients carrying this variant present a milder phenotype, as the LDLR function is not totally impaired.

In conclusion, when a variant is functionally assessed and proved to be pathogenic, is important to contextualize the results according to patients' genotype. It is imperative to understand if patients under study are homozygotes or heterozygotes, as this fact will be determinant to the aggressiveness of patient's phenotype.

It is worth mentioning that the majority of the variants are located either in the EGF precursor like domain or in the ligand-binding domain, probably because these are the biggest domains of the LDLR. From the analysis of above referred databases was possible to conclude that the majority of variants occurring in these domains, which have been

functionally studied, were proven to be pathogenic. Nevertheless, the justification for this may rely on the fact that, when a variant is proved to be neutral, these studies are not so often published.

In order to complete information here achieved, the assessment of all variants under study by confocal laser scan microscopy would allow the confirmation of results here observed and the visualization of the receptor within the cell. This would allow the characterization of variant c.1802A>T p.(Asp601Val), which could be class I, II A or II B, and variant c.1775G>A p.(Gly592Glu), to confirm its recycling defect.

Ten variants were assessed during this project, resulting in a more detailed characterization of the cause of FH in individuals carrying these variants. Functional studies are extremely necessary in case of uncertain FH diagnosis, as it happened to 97 Portuguese FH patients carrying these variants. Results here obtained allow the stratification of patients according to phenotype severity and, consequently, with cardiovascular risk, allowing a more efficient clinic and therapeutic counselling.

4.3. *In silico* vs. *in vitro*

The phenotypic impact of a missense variant depends on criteria such as the evolutionary conservation of an amino acid or nucleotide, the location and context within the protein sequence, and the biochemical consequences of the amino acid substitution [56]. These kind of parameters can be firstly assessed through the use of bioinformatics tools.

Most algorithms for missense variant prediction are 65-80% accurate when examining known disease variants [116]. Among the missense variant prediction tools, Polymorphism Phenotyping (PolyPhen-2) [64], Sorting Tolerant From Intolerant (SIFT) [65] and Mutationtaster [66] are the most commonly used for missense variant interpretation in clinical laboratories [56].

Tools to assess splice site implying variants are also available, having these higher sensitivity (~90-100%) relative to specificity (~60-80%) in predicting splice site abnormalities [117]. Among the most commonly used are the Human Splicing Finder (HSF) (http://www.umd.be/HSF3/HSF.html [67]), the Splice Site Prediction by Neutral Network (NNSSP) (http://www.fruitfly.org/seq_tools/splice.html [68]) and the FSPLICE (http://linux1.softberry.com/).

Nevertheless, these tools only provide predictions, which have an error associated, so their use in sequence variant interpretation should be implemented carefully, not being recommended as the only source of evidence to make a clinical assertion [56].

In an era where bioinformatics tools are getting better every day, one can fall in the temptation of using it as much as possible. A variety of *in silico* tools, either publicly or commercially available might be useful when interpreting variants identified in patients. These tools use different algorithms, including determination of the effect of the sequence variant several levels such as at the nucleotide, amino acid and potential impact on the protein.

Among all variants assessed by *in silico* tools (see Results, *tables 3.2.* and *3.3.*), it was possible to compare some *in silico* predictions with results previously reported and with results here obtained (see *Results, table 3.4.*). Although variants identified in *3.1. Molecular diagnosis* predictions have been 100% (7/7) correct comparing with functional assessment results, for the variants here assessed in *3.3. Functional studies*, predictions were correct only for 70% (7/10) of variants. If all variants were considered as a single group, 14/17 variants were correctly predicted, leading to an accuracy of 82%.

The effect of a missense alteration depends on criteria such as the evolutionary conservation of an amino acid or nucleotide, the location and context within the protein sequence, and the biochemical consequence of the amino acid substitution [56]. Thus, here the main purpose of using several *in silico* prediction software was to combine predictions based in distinct characteristics, in order to better assess the predicted impact of a missense alteration. [56].

According to literature, in general, most algorithms for missense variant prediction are 65–80% accurate when examining known disease variants [116], which meets the value of 82% here presented. The reason why prediction is not totally accurate may rely in the fact that most tools also tend to have low specificity, resulting in overprediction of missense alterations as deleterious [64].

Predictive *in silico* tools have potential value in disease diagnosis, in view of the impossibility of laboratory functionally testing large numbers of variants in daily clinical practice. However, it seems correct to conclude that this analysis for complex proteins, as LDLR and APOB, can reveal some limitations. This means that their use in sequence variant interpretation should be implemented carefully and should not be the single source of evidence to make a clinical assertion.

The ultimate test to confirm that a variant is disease causing is the performance of functional analysis of a variant gene, usually by DNA expression with measurement of biological activity [55]. After a variant identification and a careful analysis of the family, functional studies can be a powerful tool in support of a variant pathogenicity if they reflect closely the biological environment.

Chapter 5

Conclusion and Future perspectives

In the past 100 years scientists have made vast advances in understanding cholesterol, lipoproteins, and the mode in which genes and diets alter lipoprotein levels. Experiments using four lines of evidence— experimental, genetic, epidemiologic, and therapeutic – allowed the understanding of FH as a genetic disease for which, fortunately, therapeutic measures are available.

During this project, it was possible to observe that, although more than 1600 variants have been described in the *LDLR*, there are still novel variants being found, proving the heterogeneity of FH. It was also possible to confirm that currently available criteria for clinical diagnosis is not as specific as desirable, mainly due to environmental interaction. This led to the conclusion that the creation of clinical criteria adapted to each population characteristics, such as genetic origin, lifestyle and diet, would be ideal in the sense that no resources would be spent in molecular diagnosis of false positive FH patients. For this, new risk factors for this disease must be investigated, as well as possible new genes associated with this disorder.

In addition to novel variants, variants already reported are often found in several patients. Their impact can be predicted by *in silico* tools, however results here obtained suggest that the use of these tools should not be the only source of evidence when performing a genetic diagnosis, as it has associated errors. For this reason, *in silico* prediction does not allow a definite molecular diagnosis of FH.

Functional assessment of these variants showed that some of them might have a significant impact in the *LDLR* function. However, further studies should be performed in order to produce more lines of evidence and confirm results obtained by flow cytometry.

In the Portuguese FH Study, there are still about 25 identified without functional assessment to date, which are currently under functional study. The realization of functional studies as an integrated part of molecular diagnosis should be implemented for all cohorts, as it will contribute for the elucidation of the molecular basis of FH worldwide.

Furthermore, since a correct diagnosis is performed, early personalized counseling and treatment can be implemented, improving FH patients' prognosis and providing them a longer and better life!

Chapter 6

References

[1] Rifai N, Bachorik PS, Albers JJ. Lipids, lipoprotein and apolipoprotein. In: Burtis CA, Ashwood E.R., editors. Tietz textbook of clinical chemistry. 3rd ed. Philadelphia: W.B. Saunders company; 1999. pp. 806–61.

[2] J. L. Goldstein and M. S. Brown, "History of science. A golden era of Nobel laureates," *Science*, vol. 338, no. 6110, pp. 1033–1034, Nov. 2012.

[3] R. Zetterström, "The 1964 Nobel Prize for the discovery of the biosynthesis of cholesterol," *Acta Paediatr. Oslo Nor. 1992*, vol. 98, no. 7, pp. 1223–1227, Jul. 2009.

[4] M. A. Shampo and R. A. Kyle, "Konrad Bloch--Nobel Prize for research on cholesterol," *Mayo Clin. Proc.*, vol. 73, no. 8, p. 764, Aug. 1998.

[5] M. S. Brown and J. L. Goldstein, "A receptor-mediated pathway for cholesterol homeostasis," *Science*, vol. 232, no. 4746, pp. 34–47, Apr. 1986.

[6] A. Mazein, S. Watterson, W.-Y. Hsieh, W. J. Griffiths, and P. Ghazal, "A comprehensive machine-readable view of the mammalian cholesterol biosynthesis pathway," *Biochem. Pharmacol.*, vol. 86, no. 1, pp. 56–66, Jul. 2013.

[7] J. L. Goldstein and M. S. Brown, "A century of cholesterol and coronaries: from plaques to genes to statins," *Cell*, vol. 161, no. 1, pp. 161–172, Mar. 2015.

[8] O. F. De Lalla and J. W. Gofman, "Ultracentrifugal analysis of serum lipoproteins," *Methods Biochem. Anal.*, vol. 1, pp. 459–478, 1954.

[9] J. L. Oncley, "Lipoproteins of human plasma," *Harvey Lect.*, vol. 50, pp. 71–91, 1955 1954.

[10] D. S. Fredrickson, "Plasma lipoproteins and apolipoproteins," *Harvey Lect.*, vol. 68, pp. 185–237, 1974.

[11] R. E. Olson, "Discovery of the Lipoproteins, Their Role in Fat Transport and Their Significance as Risk Factors," *J. Nutr.*, vol. 128, no. 2, p. 439S–443S, Feb. 1998.

[12] R. A. Cox and M. R. García-Palmieri, "Cholesterol, Triglycerides, and Associated Lipoproteins," in *Clinical Methods: The History, Physical, and Laboratory Examinations*, 3rd ed., H. K. Walker, W. D. Hall, and J. W. Hurst, Eds. Boston: Butterworths, 1990.

[13] L. Mateu, A. Tardieu, V. Luzzati, L. Aggerbeck, and A. M. Scanu, "On the structure of human serum low density lipoprotein," *J. Mol. Biol.*, vol. 70, no. 1, pp. 105–116, Sep. 1972.

[14] J. D. Morrisett, R. L. Jackson, and A. M. Gotto, "Lipoproteins: Structure and Function," *Annu. Rev. Biochem.*, vol. 44, no. 1, pp. 183–207, 1975.

[15] H. H. Hobbs, M. S. Brown, and J. L. Goldstein, "Molecular genetics of the LDL receptor gene in familial hypercholesterolemia," *Hum. Mutat.*, vol. 1, no. 6, pp. 445–466, 1992.

[16] E. R. Eden, X.-M. Sun, D. D. Patel, and A. K. Soutar, "Adaptor protein disabled-2 modulates low density lipoprotein receptor synthesis in fibroblasts from patients with autosomal recessive hypercholesterolaemia," *Hum. Mol. Genet.*, vol. 16, no. 22, pp. 2751–2759, Nov. 2007.

[17] P. Michaely, W.-P. Li, R. G. W. Anderson, J. C. Cohen, and H. H. Hobbs, "The modular adaptor protein ARH is required for low density lipoprotein (LDL) binding and internalization but not for LDL receptor clustering in coated pits," *J. Biol. Chem.*, vol. 279, no. 32, pp. 34023–34031, Aug. 2004.

[18] C. G. Davis, J. L. Goldstein, T. C. Südhof, R. G. Anderson, D. W. Russell, and M. S. Brown, "Acid-dependent ligand dissociation and recycling of LDL receptor mediated by growth factor homology region," *Nature*, vol. 326, no. 6115, pp. 760–765, Apr. 1987.

[19] M. Abifadel, J.-P. Rabès, M. Devillers, A. Munnich, D. Erlich, C. Junien, M. Varret, and C. Boileau, "Mutations and polymorphisms in the proprotein convertase subtilisin kexin 9 (PCSK9) gene in cholesterol metabolism and disease," *Hum. Mutat.*, vol. 30, no. 4, pp. 520–529, Apr. 2009.

[20] G.-W. Go and A. Mani, "Low-density lipoprotein receptor (LDLR) family orchestrates cholesterol homeostasis," *Yale J. Biol. Med.*, vol. 85, no. 1, pp. 19–28, Mar. 2012.

[21] G. K. Hovingh, M. H. Davidson, J. J. P. Kastelein, and A. M. O'Connor, "Diagnosis and treatment of familial hypercholesterolaemia," *Eur. Heart J.*, vol. 34, no. 13, pp. 962–971, Apr. 2013.

[22] C. Müller, "Xanthomata, Hypercholesterolemia, Angina Pectoris.," *Acta Med. Scand.*, vol. 95, no. S89, pp. 75–84, Jan. 1938.

[23] B. G. Nordestgaard, M. J. Chapman, S. E. Humphries, H. N. Ginsberg, L. Masana, O. S. Descamps, O. Wiklund, R. A. Hegele, F. J. Raal, J. C. Defesche, A. Wiegman, R. D. Santos, G. F. Watts, K. G. Parhofer, G. K. Hovingh, P. T. Kovanen, C. Boileau, M. Averna, J. Borén, E. Bruckert, A. L. Catapano, J. A. Kuivenhoven, P. Pajukanta, K. Ray, A. F. H. Stalenhoef, E. Stroes, M.-R. Taskinen, A. Tybjærg-Hansen, and European Atherosclerosis Society Consensus Panel, "Familial hypercholesterolaemia is underdiagnosed and undertreated in the general population: guidance for clinicians to prevent coronary heart disease: consensus statement of the European Atherosclerosis Society," *Eur. Heart J.*, vol. 34, no. 45, pp. 3478–3490a, Dec. 2013.

[24] S. Singh and V. Bittner, "Familial hypercholesterolemia--epidemiology, diagnosis, and screening," *Curr. Atheroscler. Rep.*, vol. 17, no. 2, p. 482, 2015.

[25] J. A. Marbach, J. L. McKeon, J. L. Ross, and D. Duffy, "Novel treatments for familial hypercholesterolemia: pharmacogenetics at work," *Pharmacotherapy*, vol. 34, no. 9, pp. 961–972, Sep. 2014.

[26] A. Wiegman, S. S. Gidding, G. F. Watts, M. J. Chapman, H. N. Ginsberg, M. Cuchel, L. Ose, M. Averna, C. Boileau, J. Borén, E. Bruckert, A. L. Catapano, J. C. Defesche, O. S. Descamps, R. A. Hegele, G. K. Hovingh, S. E. Humphries, P. T. Kovanen, J. A. Kuivenhoven, L. Masana, B. G. Nordestgaard, P. Pajukanta, K. G. Parhofer, F. J. Raal, K. K. Ray, R. D. Santos, A. F. H. Stalenhoef, E. Steinhagen-Thiessen, E. S. Stroes, M.-R. Taskinen, A. Tybjærg-Hansen, O. Wiklund, and European Atherosclerosis Society Consensus Panel, "Familial hypercholesterolaemia in children and adolescents: gaining decades of life by optimizing detection and treatment," *Eur. Heart J.*, vol. 36, no. 36, pp. 2425–2437, Sep. 2015.

[27] A. Endo, M. Kuroda, and K. Tanzawa, "Competitive inhibition of 3-hydroxy-3-methylglutaryl coenzyme A reductase by ML-236A and ML-236B fungal metabolites, having hypocholesterolemic activity," *FEBS Lett.*, vol. 72, no. 2, pp. 323–326, Dec. 1976.

[28] A. M. Gotto, "Statins: Powerful Drugs for Lowering Cholesterol Advice for Patients," *Circulation*, vol. 105, no. 13, pp. 1514–1516, Apr. 2002.

[29] R. Uauy, G. L. Vega, S. M. Grundy, and D. M. Bilheimer, "Lovastatin therapy in receptor-negative homozygous familial hypercholesterolemia: lack of effect on low-density lipoprotein concentrations or turnover," *J. Pediatr.*, vol. 113, no. 2, pp. 387–392, Aug. 1988.

[30] C. Gagné, D. Gaudet, E. Bruckert, and for the E. S. Group, "Efficacy and Safety of Ezetimibe Coadministered With Atorvastatin or Simvastatin in Patients With Homozygous Familial Hypercholesterolemia," *Circulation*, vol. 105, no. 21, pp. 2469–2475, May 2002.

[31] D. J. Rader, J. Cohen, and H. H. Hobbs, "Monogenic hypercholesterolemia: new insights in pathogenesis and treatment," *J. Clin. Invest.*, vol. 111, no. 12, pp. 1795–1803, Jun. 2003.

[32] S. J. Bandeali, J. Daye, and S. S. Virani, "Novel therapies for treating familial hypercholesterolemia," *Curr. Atheroscler. Rep.*, vol. 16, no. 1, p. 382, Jan. 2014.

[33] G. R. Thompson, M. Barbir, D. Davies, P. Dobral, M. Gesinde, M. Livingston, P. Mandry, A. D. Marais, S. Matthews, C. Neuwirth, A. Pottle, C. le Roux, D. Scullard, C. Tyler, and S. Watkins, "Efficacy criteria and cholesterol targets for LDL apheresis," *Atherosclerosis*, vol. 208, no. 2, pp. 317–321, Feb. 2010.

[34] M. P. McGowan, "Emerging low-density lipoprotein (LDL) therapies: Management of severely elevated LDL cholesterol--the role of LDL-apheresis," *J. Clin. Lipidol.*, vol. 7, no. 3 Suppl, pp. S21–26, Jun. 2013.

[35] S. G. Young and L. G. Fong, "Lowering plasma cholesterol by raising LDL receptors--revisited," *N. Engl. J. Med.*, vol. 366, no. 12, pp. 1154–1155, Mar. 2012.

[36] R. A. Vogel, "PCSK9 inhibition: the next statin?," *J. Am. Coll. Cardiol.*, vol. 59, no. 25, pp. 2354–2355, Jun. 2012.

[37] T. L. Innerarity, K. H. Weisgraber, K. S. Arnold, R. W. Mahley, R. M. Krauss, G. L. Vega, and S. M. Grundy, "Familial defective apolipoprotein B-100: low density lipoproteins with abnormal receptor binding," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 84, no. 19, pp. 6919–6923, Oct. 1987.

[38] M. Abifadel, M. Varret, J.-P. Rabès, D. Allard, K. Ouguerram, M. Devillers, C. Cruaud, S. Benjannet, L. Wickham, D. Erlich, A. Derré, L. Villéger, M. Farnier, I. Beucler, E. Bruckert, J. Chambaz, B. Chanu, J.-M. Lecerf, G. Luc, P. Moulin, J. Weissenbach, A. Prat, M. Krempf, C. Junien, N. G. Seidah, and C. Boileau, "Mutations in PCSK9 cause autosomal dominant hypercholesterolemia," *Nat. Genet.*, vol. 34, no. 2, pp. 154–156, Jun. 2003.

[39] M. A. Austin, C. M. Hutter, R. L. Zimmern, and S. E. Humphries, "Genetic Causes of Monogenic Heterozygous Familial Hypercholesterolemia: A HuGE Prevalence Review," *Am. J. Epidemiol.*, vol. 160, no. 5, pp. 407–420, Sep. 2004.

[40] A. K. Soutar and R. P. Naoumova, "Mechanisms of disease: genetic causes of familial hypercholesterolemia," *Nat. Clin. Pract. Cardiovasc. Med.*, vol. 4, no. 4, pp. 214–225, Apr. 2007.

[41] A. C. Alves, A. Etxebarria, A. K. Soutar, C. Martin, and M. Bourbon, "Novel functional APOB mutations outside LDL-binding region causing familial hypercholesterolaemia," *Hum. Mol. Genet.*, vol. 23, no. 7, pp. 1817–1828, Apr. 2014.

[42] T. A. Lagace, D. E. Curtis, R. Garuti, M. C. McNutt, S. W. Park, H. B. Prather, N. N. Anderson, Y. K. Ho, R. E. Hammer, and J. D. Horton, "Secreted PCSK9 decreases the number of LDL receptors in hepatocytes and in livers of parabiotic mice," *J. Clin. Invest.*, vol. 116, no. 11, pp. 2995–3005, Nov. 2006.

[43] D.-W. Zhang, T. A. Lagace, R. Garuti, Z. Zhao, M. McDonald, J. D. Horton, J. C. Cohen, and H. H. Hobbs, "Binding of proprotein convertase subtilisin/kexin type 9 to epidermal growth factor-like repeat A of low density lipoprotein receptor decreases receptor recycling and increases degradation," *J. Biol. Chem.*, vol. 282, no. 25, pp. 18602–18612, Jun. 2007.

[44] J. C. Cohen, E. Boerwinkle, T. H. Mosley, and H. H. Hobbs, "Sequence variations in PCSK9, low LDL, and protection against coronary heart disease," *N. Engl. J. Med.*, vol. 354, no. 12, pp. 1264–1272, Mar. 2006.

[45] M. Abifadel, S. Elbitar, P. El Khoury, Y. Ghaleb, M. Chémaly, M.-L. Moussalli, J.-P. Rabès, M. Varret, and C. Boileau, "Living the PCSK9 adventure: from the identification of a new gene in familial hypercholesterolemia towards a potential new class of anticholesterol drugs," *Curr. Atheroscler. Rep.*, vol. 16, no. 9, p. 439, Sep. 2014.

[46] C. K. Garcia, K. Wilund, M. Arca, G. Zuliani, R. Fellin, M. Maioli, S. Calandra, S. Bertolini, F. Cossu, N. Grishin, R. Barnes, J. C. Cohen, and H. H. Hobbs, "Autosomal Recessive Hypercholesterolemia Caused by Mutations in a Putative LDL Receptor Adaptor Protein," *Science*, vol. 292, no. 5520, pp. 1394–1398, May 2001.

[47] R. Garuti, C. Jones, W.-P. Li, P. Michaely, J. Herz, R. D. Gerard, J. C. Cohen, and H. H. Hobbs, "The modular adaptor protein autosomal recessive hypercholesterolemia (ARH) promotes low density lipoprotein receptor clustering into clathrin-coated pits," *J. Biol. Chem.*, vol. 280, no. 49, pp. 40996–41004, Dec. 2005.

[48] L. Pisciotta, C. Priore Oliva, G. M. Pes, L. Di Scala, A. Bellocchio, R. Fresa, A. Cantafora, M. Arca, S. Calandra, and S. Bertolini, "Autosomal recessive hypercholesterolemia (ARH) and homozygous familial hypercholesterolemia (FH): a phenotypic comparison," *Atherosclerosis*, vol. 188, no. 2, pp. 398–405, Oct. 2006.

[49] T. C. Südhof, D. W. Russell, J. L. Goldstein, M. S. Brown, R. Sanchez-Pescador, and G. I. Bell, "Cassette of eight exons shared by genes for LDL receptor and EGF precursor," *Science*, vol. 228, no. 4701, pp. 893–895, May 1985.

[50] D. W. Russell, W. J. Schneider, T. Yamamoto, K. L. Luskey, M. S. Brown, and J. L. Goldstein, "Domain map of the LDL receptor: sequence homology with the epidermal growth factor precursor," *Cell*, vol. 37, no. 2, pp. 577–585, Jun. 1984.

[51] T. Yamamoto, C. G. Davis, M. S. Brown, W. J. Schneider, M. L. Casey, J. L. Goldstein, and D. W. Russell, "The human LDL receptor: a cysteine-rich protein with multiple Alu sequences in its mRNA," *Cell*, vol. 39, no. 1, pp. 27–38, Nov. 1984.

[52] Goldstein JL, Hobbs HH, Brown MS. In: The Metabolic and Molecular Basis of Inherited Diseases. Scriver CR, Beaudet AL, Sly WS, Valle D, Childs B, Kinzler KW, Vogelstien B, editor. New York: McGraw-Hill; 2001. Familial hypercholesterolaemia; pp. 2863–2913

[53] J. L. Goldstein, M. S. Brown, R. G. Anderson, D. W. Russell, and W. J. Schneider, "Receptor-mediated endocytosis: concepts emerging from the LDL receptor system," *Annu. Rev. Cell Biol.*, vol. 1, pp. 1–39, 1985.

[54] T. B. Strøm, J. K. Laerdahl, and T. P. Leren, "Mutation p.L799R in the LDLR, which affects the transmembrane domain of the LDLR, prevents membrane insertion and causes secretion of the mutant LDLR," *Hum. Mol. Genet.*, Jul. 2015.

[55] R. G. Cotton and C. R. Scriver, "Proof of 'disease causing' mutation," *Hum. Mutat.*, vol. 12, no. 1, pp. 1–3, 1998.

[56] S. Richards, N. Aziz, S. Bale, D. Bick, S. Das, J. Gastier-Foster, W. W. Grody, M. Hegde, E. Lyon, E. Spector, K. Voelkerding, H. L. Rehm, and ACMG Laboratory Quality Assurance Committee, "Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology," *Genet. Med. Off. J. Am. Coll. Med. Genet.*, vol. 17, no. 5, pp. 405–424, May 2015.

[57] M. Bourbon, Q. Rato, and Investigadores do Estudo Português de Hipercolesterolemia Familiar, "Portuguese Familial Hypercholesterolemia Study: presentation of the study and preliminary results," *Rev. Port. Cardiol. Orgão Of. Soc. Port. Cardiol. Port. J. Cardiol. Off. J. Port. Soc. Cardiol.*, vol. 25, no. 11, pp. 999–1013, Nov. 2006.

[58] A. M. Medeiros, A. C. Alves, and M. Bourbon, "Mutational analysis of a cohort with clinical diagnosis of familial hypercholesterolemia: considerations for genetic diagnosis improvement," *Genet. Med. Off. J. Am. Coll. Med. Genet.*, May 2015.

[59] G. F. Watts, S. Gidding, A. S. Wierzbicki, P. P. Toth, R. Alonso, W. V. Brown, E. Bruckert, J. Defesche, K. K. Lin, M. Livingston, P. Mata, K. G. Parhofer, F. J. Raal, R. D. Santos, E. J. G. Sijbrands, W. G. Simpson, D. R. Sullivan, A. V. Susekov, B. Tomlinson, A. Wiegman, S. Yamashita, and J. J. P. Kastelein, "Integrated guidance on the care of familial hypercholesterolaemia from the International FH Foundation," *Eur. J. Prev. Cardiol.*, vol. 22, no. 7, pp. 849–854, Jul. 2015.

[60] R. Alonso, P. Mata, D. Zambón, N. Mata, and F. Fuentes-Jiménez, "Early diagnosis and treatment of familial hypercholesterolemia: improving patient outcomes," *Expert Rev. Cardiovasc. Ther.*, vol. 11, no. 3, pp. 327–342, Mar. 2013.

[61] "Risk of fatal coronary heart disease in familial hypercholesterolaemia. Scientific Steering Committee on behalf of the Simon Broome Register Group.," *BMJ*, vol. 303, no. 6807, pp. 893–896, Oct. 1991.

[62] D. K. Lahiri and J. I. Nurnberger, "A rapid non-enzymatic method for the preparation of HMW DNA from blood for RFLP studies," *Nucleic Acids Res.*, vol. 19, no. 19, p. 5444, Oct. 1991.

[63] J. T. den Dunnen and S. E. Antonarakis, "Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion," *Hum. Mutat.*, vol. 15, no. 1, pp. 7–12, 2000.

[64] I. Adzhubei, D. M. Jordan, and S. R. Sunyaev, "Predicting functional effect of human missense mutations using PolyPhen-2," *Curr. Protoc. Hum. Genet. Editor. Board Jonathan Haines Al*, vol. Chapter 7, p. Unit7.20, Jan. 2013.

[65] P. C. Ng and S. Henikoff, "SIFT: Predicting amino acid changes that affect protein function," *Nucleic Acids Res.*, vol. 31, no. 13, pp. 3812–3814, Jul. 2003.

[66] J. M. Schwarz, C. Rödelsperger, M. Schuelke, and D. Seelow, "MutationTaster evaluates disease-causing potential of sequence alterations," *Nat. Methods*, vol. 7, no. 8, pp. 575–576, Aug. 2010.

[67] F.-O. Desmet, D. Hamroun, M. Lalande, G. Collod-Béroud, M. Claustres, and C. Béroud, "Human Splicing Finder: an online bioinformatics tool to predict splicing signals," *Nucleic Acids Res.*, vol. 37, no. 9, p. e67, May 2009.

[68] A. A. Salamov and V. V. Solovyev, "Prediction of protein secondary structure by combining nearest-neighbor algorithms and multiple sequence alignments," *J. Mol. Biol.*, vol. 247, no. 1, pp. 11–15, Mar. 1995.

[69] A. Etxebarria, A. Benito-Vicente, A. C. Alves, H. Ostolaza, M. Bourbon, and C. Martin, "Advantages and versatility of fluorescence-based methodology to characterize the functionality of LDLR and class mutation assignment," *PloS One*, vol. 9, no. 11, p. e112677, 2014.

[70] M. J. Chapman, S. Goldstein, D. Lagrange, and P. M. Laplaud, "A density gradient ultracentrifugal procedure for the isolation of the major lipoprotein classes from human serum," *J. Lipid Res.*, vol. 22, no. 2, pp. 339–358, Feb. 1981.

[71] J. Picot, C. L. Guerin, C. Le Van Kim, and C. M. Boulanger, "Flow cytometry: retrospective, fundamentals and recent instrumentation," *Cytotechnology*, vol. 64, no. 2, pp. 109–130, Mar. 2012.

[72] M. Bourbon, A. C. Alves, A. M. Medeiros, S. Silva, A. K. Soutar, and Investigators of Portuguese FH Study, "Familial hypercholesterolaemia in Portugal," *Atherosclerosis*, vol. 196, no. 2, pp. 633–642, Feb. 2008.

[73] A. Benito-Vicente, A. C. Alves, A. Etxebarria, A. M. Medeiros, C. Martin, and M. Bourbon, "The importance of an integrated analysis of clinical, molecular, and functional data for the genetic diagnosis of familial hypercholesterolemia," *Genet. Med. Off. J. Am. Coll. Med. Genet.*, Mar. 2015.

[74] A. M. Medeiros, A. C. Alves, V. Francisco, M. Bourbon, and investigators of the Portuguese FH Study, "Update of the Portuguese Familial Hypercholesterolaemia Study," *Atherosclerosis*, vol. 212, no. 2, pp. 553–558, Oct. 2010.

[75] L. F. Soria, E. H. Ludwig, H. R. Clarke, G. L. Vega, S. M. Grundy, and B. J. McCarthy, "Association between a specific apolipoprotein B mutation and familial defective apolipoprotein B-100," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 86, no. 2, pp. 587–591, Jan. 1989.

[76] T. L. Innerarity, R. W. Mahley, K. H. Weisgraber, T. P. Bersot, R. M. Krauss, G. L. Vega, S. M. Grundy, W. Friedl, J. Davignon, and B. J. McCarthy, "Familial defective apolipoprotein B-100: a mutation of apolipoprotein B that causes hypercholesterolemia," *J. Lipid Res.*, vol. 31, no. 8, pp. 1337–1349, Aug. 1990.

[77] J. C. Webb, X. M. Sun, S. N. McCarthy, C. Neuwirth, G. R. Thompson, B. L. Knight, and A. K. Soutar, "Characterization of mutations in the low density lipoprotein (LDL)-receptor gene in patients with homozygous familial hypercholesterolemia, and frequency of these mutations in FH patients in the United Kingdom," *J. Lipid Res.*, vol. 37, no. 2, pp. 368–381, Feb. 1996.

[78] T. P. Leren, K. Solberg, O. K. Rødningen, S. Tonstad, and L. Ose, "Two founder mutations in the LDL receptor gene in Norwegian familial hypercholesterolemia subjects," *Atherosclerosis*, vol. 111, no. 2, pp. 175–182, Dec. 1994.

[79] L. G. Jensen, H. K. Jensen, H. Nissen, K. Kristiansen, O. Faergeman, L. Bolund, and N. Gregersen, "An LDL receptor promoter mutation in a heterozygous FH patient with dramatically skewed ratio between the two allelic mRNA variants," *Hum. Mutat.*, vol. 7, no. 1, pp. 82–84, 1996.

[80] J. Cameron, Ø. L. Holla, M. A. Kulseth, T. P. Leren, and K. E. Berge, "Splice-site mutation c.313+1, G>A in intron 3 of the LDL receptor gene results in transcripts with skipping of exon 3 and inclusion of intron 3," *Clin. Chim. Acta Int. J. Clin. Chem.*, vol. 403, no. 1–2, pp. 131–135, May 2009.

[81] J. C. Webb, X. M. Sun, D. D. Patel, S. N. McCarthy, B. L. Knight, and A. K. Soutar, "Characterization of two new point mutations in the low density lipoprotein receptor genes of an English patient with homozygous familial hypercholesterolemia," *J. Lipid Res.*, vol. 33, no. 5, pp. 689–698, May 1992.

[82] W. K. Lee, L. Haddad, M. J. Macleod, A. M. Dorrance, D. J. Wilson, D. Gaffney, M. H. Dominiczak, C. J. Packard, I. N. Day, S. E. Humphries, and A. F. Dominiczak, "Identification of a common low density lipoprotein receptor mutation (C163Y) in the west of Scotland," *J. Med. Genet.*, vol. 35, no. 7, pp. 573–578, Jul. 1998.

[83] X. M. Sun, D. D. Patel, B. L. Knight, and A. K. Soutar, "Comparison of the genetic defect with LDL-receptor activity in cultured cells from patients with a clinical diagnosis of heterozygous familial hypercholesterolemia. The Familial Hypercholesterolaemia Regression Study Group," *Arterioscler. Thromb. Vasc. Biol.*, vol. 17, no. 11, pp. 3092–3101, Nov. 1997.

[84] S. W. Fouchier, J. J. P. Kastelein, and J. C. Defesche, "Update of the molecular basis of familial hypercholesterolemia in The Netherlands," *Hum. Mutat.*, vol. 26, no. 6, pp. 550–556, Dec. 2005.

[85] P. Mozas, S. Castillo, D. Tejedor, G. Reyes, R. Alonso, M. Franco, P. Saenz, F. Fuentes, F. Almagro, P. Mata, and M. Pocoví, "Molecular characterization of familial hypercholesterolemia in Spain: identification of 39 novel and 77 recurrent mutations in LDLR," *Hum. Mutat.*, vol. 24, no. 2, p. 187, Aug. 2004.

[86] U. M. Koivisto, J. S. Viikari, and K. Kontula, "Molecular characterization of minor gene rearrangements in Finnish patients with heterozygous familial hypercholesterolemia: identification of two common missense mutations (Gly823-->Asp and Leu380-->His) and eight rare mutations of the LDL receptor gene," *Am. J. Hum. Genet.*, vol. 57, no. 4, pp. 789–797, Oct. 1995.

[87] A. M. Medeiros, A. C. Alves, P. Aguiar, M. Bourbon, and Pediatric Investigators of the Portuguese Familial Hypercholesterolemia Study, "Cardiovascular risk assessment of dyslipidemic children: analysis of biomarkers to identify monogenic dyslipidemia," *J. Lipid Res.*, vol. 55, no. 5, pp. 947–955, May 2014.

[88] P. N. Hopkins, P. P. Toth, C. M. Ballantyne, D. J. Rader, and National Lipid Association Expert Panel on Familial Hypercholesterolemia, "Familial hypercholesterolemias: prevalence, genetics, diagnosis and screening recommendations from the National Lipid Association Expert Panel on Familial Hypercholesterolemia," *J. Clin. Lipidol.*, vol. 5, no. 3 Suppl, pp. S9–17, Jun. 2011.

[89] S. R. Daniels, F. R. Greer, and Committee on Nutrition, "Lipid screening and cardiovascular health in childhood," *Pediatrics*, vol. 122, no. 1, pp. 198–208, Jul. 2008.

[90] O. Guardamagna, G. Restagno, E. Rolfo, C. Pederiva, S. Martini, F. Abello, V. Baracco, L. Pisciotta, E. Pino, S. Calandra, and S. Bertolini, "The type of LDLR gene mutation predicts cardiovascular risk in children with familial hypercholesterolemia," *J. Pediatr.*, vol. 155, no. 2, pp. 199–204.e2, Aug. 2009.

[91] M. Marduel, K. Ouguerram, V. Serre, D. Bonnefont-Rousselot, A. Marques-Pinheiro, K. Erik Berge, M. Devillers, G. Luc, J.-M. Lecerf, L. Tosolini, D. Erlich, G. M. Peloso, N. Stitziel, P. Nitchké, J.-P. Jaïs, French Research Network on ADH, M. Abifadel, S. Kathiresan, T. P. Leren, J.-P. Rabès, C. Boileau, and M. Varret, "Description of a large family with autosomal dominant hypercholesterolemia associated with the APOE p.Leu167del mutation," *Hum. Mutat.*, vol. 34, no. 1, pp. 83–87, Jan. 2013.

[92] B. Martín, M. Solanas-Barca, A.-L. García-Otín, S. Pampín, M. Cofán, E. Ros, J.-C. Rodríguez-Rey, M. Pocoví, and F. Civeira, "An NPC1L1 gene promoter variant is associated with autosomal dominant hypercholesterolemia," *Nutr. Metab. Cardiovasc. Dis. NMCD*, vol. 20, no. 4, pp. 236–242, May 2010.

[93] J. de Graaf, G. van der Vleuten, and A. F. H. Stalenhoef, "Diagnostic criteria in relation to the pathogenesis of familial combined hyperlipidemia," *Semin. Vasc. Med.*, vol. 4, no. 3, pp. 229–240, Aug. 2004.

[94] B. G. Nordestgaard, M. J. Chapman, K. Ray, J. Borén, F. Andreotti, G. F. Watts, H. Ginsberg, P. Amarenco, A. Catapano, O. S. Descamps, E. Fisher, P. T. Kovanen, J. A. Kuivenhoven, P. Lesnik, L. Masana, Z. Reiner, M.-R. Taskinen, L. Tokgözoglu, A. Tybjærg-Hansen, and European Atherosclerosis Society Consensus Panel, "Lipoprotein(a) as a cardiovascular risk factor: current status," *Eur. Heart J.*, vol. 31, no. 23, pp. 2844–2853, Dec. 2010.

[95] R. Romagnuolo, C. A. Scipione, M. B. Boffa, S. M. Marcovina, N. G. Seidah, and M. L. Koschinsky, "Lipoprotein(a) catabolism is regulated by proprotein convertase subtilisin/kexin type 9 through the low density lipoprotein receptor," *J. Biol. Chem.*, vol. 290, no. 18, pp. 11649–11662, May 2015.

[96] D. Marks, D. Wonderling, M. Thorogood, H. Lambert, S. E. Humphries, and H. A. W. Neil, "Cost effectiveness analysis of different approaches of screening for familial hypercholesterolaemia," *BMJ*, vol. 324, no. 7349, p. 1303, Jun. 2002.

[97] T. P. Leren, S. Tonstad, K. E. Gundersen, K. S. Bakken, O. K. Rødningen, H. Sundvold, L. Ose, and K. Berg, "Molecular genetics of familial hypercholesterolaemia in Norway," *J. Intern. Med.*, vol. 241, no. 3, pp. 185–194, Mar. 1997.

[98] I. F. A. C. Fokkema, P. E. M. Taschner, G. C. P. Schaafsma, J. Celli, J. F. J. Laros, and J. T. den Dunnen, "LOVD v.2.0: the next generation in gene variant databases," *Hum. Mutat.*, vol. 32, no. 5, pp. 557–563, May 2011.

[99] M. S. Brown, P. T. Kovanen, and J. L. Goldstein, "Evolution of the Ldl Receptor Concept–from Cultured Cells to Intact Animals *," *Ann. N. Y. Acad. Sci.*, vol. 348, no. 1, pp. 48–68, Jun. 1980.

[100] J. L. Goldstein and M. S. Brown, "Binding and degradation of low density lipoproteins by cultured human fibroblasts. Comparison of cells from a normal subject and from a patient with homozygous familial hypercholesterolemia," *J. Biol. Chem.*, vol. 249, no. 16, pp. 5153–5162, Aug. 1974.

[101] E. Leitersdorf, H. H. Hobbs, A. M. Fourie, M. Jacobs, D. R. van der Westhuyzen, and G. A. Coetzee, "Deletion in the first cysteine-rich repeat of low density lipoprotein receptor impairs its transport but not lipoprotein binding in fibroblasts from a subject with familial hypercholesterolemia," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 85, no. 21, pp. 7912–7916, Nov. 1988.

[102] G. Schmitz, T. Brüning, E. Kovacs, and S. Barlage, "Fluorescence flow cytometry of human leukocytes in the detection of LDL receptor defects in the differential diagnosis of hypercholesterolemia," *Arterioscler. Thromb. J. Vasc. Biol. Am. Heart Assoc.*, vol. 13, no. 7, pp. 1053–1065, Jul. 1993.

[103] T. Yamamoto, R. W. Bishop, M. S. Brown, J. L. Goldstein, and D. W. Russell, "Deletion in cysteine-rich region of LDL receptor impedes transport to cell surface in WHHL rabbit," *Science*, vol. 232, no. 4755, pp. 1230–1237, Jun. 1986.

[104] M. S. Brown and J. L. Goldstein, "Regulation of the activity of the low density lipoprotein receptor in human fibroblasts," *Cell*, vol. 6, no. 3, pp. 307–316, Nov. 1975.

[105] T. Ranheim, M. A. Kulseth, K. E. Berge, and T. P. Leren, "Model system for phenotypic characterization of sequence variations in the LDL receptor gene," *Clin. Chem.*, vol. 52, no. 8, pp. 1469–1479, Aug. 2006.

[106] A. Etxebarria, L. Palacios, M. Stef, D. Tejedor, K. B. Uribe, A. Oleaga, L. Irigoyen, B. Torres, H. Ostolaza, and C. Martin, "Functional characterization of splicing and ligand-binding domain variants in the LDL receptor," *Hum. Mutat.*, vol. 33, no. 1, pp. 232–243, Jan. 2012.

[107] S. Silva, A. C. Alves, D. Patel, R. Malhó, A. K. Soutar, and M. Bourbon, "*In vitro* functional characterization of missense mutations in the LDLR gene," *Atherosclerosis*, vol. 225, no. 1, pp. 128–134, Nov. 2012.

[108] J. Cline, J. C. Braman, and H. H. Hogrefe, "PCR fidelity of pfu DNA polymerase and other thermostable DNA polymerases," *Nucleic Acids Res.*, vol. 24, no. 18, pp. 3546–3551, Sep. 1996.

[109] P. McInerney, P. Adams, and M. Z. Hadi, "Error Rate Comparison during Polymerase Chain Reaction by DNA Polymerase," *Mol. Biol. Int.*, vol. 2014, p. 287430, 2014.

[110] P.-C. Hsieh and R. Vaisvila, "Protein engineering: single or multiple site-directed mutagenesis," *Methods Mol. Biol. Clifton NJ*, vol. 978, pp. 173–186, 2013.

[111] I. F. A. C. Fokkema, J. T. den Dunnen, and P. E. M. Taschner, "LOVD: easy creation of a locus-specific sequence variation database using an 'LSDB-in-a-box' approach," *Hum. Mutat.*, vol. 26, no. 2, pp. 63–68, Aug. 2005.

[112] S. E. A. Leigh, A. H. Foster, R. A. Whittall, C. S. Hubbart, and S. E. Humphries, "Update and analysis of the University College London low density lipoprotein receptor familial hypercholesterolemia database," *Ann. Hum. Genet.*, vol. 72, no. Pt 4, pp. 485–498, Jul. 2008.

[113] P. D. Stenson, E. V. Ball, M. Mort, A. D. Phillips, J. A. Shiel, N. S. T. Thomas, S. Abeysinghe, M. Krawczak, and D. N. Cooper, "Human Gene Mutation Database (HGMD): 2003 update," *Hum. Mutat.*, vol. 21, no. 6, pp. 577–581, Jun. 2003.

[114] P. Lombardi, E. J. Sijbrands, K. van de Giessen, A. H. Smelt, J. J. Kastelein, R. R. Frants, and L. M. Havekes, "Mutations in the low density lipoprotein receptor gene of familial hypercholesterolemic patients detected by denaturing gradient gel electrophoresis and direct sequencing," *J. Lipid Res.*, vol. 36, no. 4, pp. 860–867, Apr. 1995.

[115] I. Mikhailenko, W. Considine, K. M. Argraves, D. Loukinov, B. T. Hyman, and D. K. Strickland, "Functional domains of the very low density lipoprotein receptor: molecular analysis of ligand binding and acid-dependent ligand dissociation mechanisms," *J. Cell Sci.*, vol. 112 (Pt 19), pp. 3269–3281, Oct. 1999.

[116] J. Thusberg, A. Olatubosun, and M. Vihinen, "Performance of mutation pathogenicity prediction methods on missense variants," *Hum. Mutat.*, vol. 32, no. 4, pp. 358–368, Apr. 2011.

[117] C. Houdayer, V. Caux-Moncoutier, S. Krieger, M. Barrois, F. Bonnet, V. Bourdon, M. Bronner, M. Buisson, F. Coulet, P. Gaildrat, C. Lefol, M. Léone, S. Mazoyer, D. Muller, A. Remenieras, F. Révillion, E. Rouleau, J. Sokolowska, J.-P. Vert, R. Lidereau, F. Soubrier, H. Sobol, N. Sevenet, B. Bressac-de Paillerets, A. Hardouin, M. Tosi, O. M. Sinilnikova, and D. Stoppa-Lyonnet, "Guidelines for splicing analysis in molecular diagnosis derived from a set of 327 combined *in silico*/*in vitro* studies on BRCA1 and BRCA2 variants," *Hum. Mutat.*, vol. 33, no. 8, pp. 1228–1238, Aug. 2012.

Appendix I - Molecular Biology Techniques

Table A I.1. *Proportion of reagents per mL of blood for genomic DNA extraction*

Blood	TKM X100	IGEPAL	TKM1	TKM2	SDS	NaCl	EtOH
1mL	1mL	25µL	1mL	160µL	10µL	60µL	460µL

Table A I.2. *PCR and Automated Sequencing – primers and annealing temperatures*

Exon	5' -3' sequence	Amplified region (bp)	Annealing Temperature	PCR Primers	Sequencing Primer
Promoter + Exon 1	F: ACAAATCAAGTCGCCTGCCC R: GCCATTACCCCACAAGTCTCC MB257: GGGTTAAAAAGCCGATGTCA	480	59º	SPr+1F SPr+1R	MB257
Exon 2	F: TCCCATACCCCAGAGAGTCCATA R: CAGCCGCCATCATCAAAAAG	587	58º	EX2F EX2R	EX2F
Exon 3	F: GGTTTCACTATATTGGCCAGG R: CTCCCCAGGACTCAGATAGG	327	59º	EX3F EX3R	EX3F
Exon 4	F: GTACAGATGAGGAAACTGAG R: TTGGCATGTTGTTGGAAATCC R4F: GAGGAAACTGAGGCACCGAG	677	57º	EX4F EX4R	EX4F
Exon 5	F: GCAAAAGGCCCTGCTTCTTT R:GAGGCTCTGAGAAGTCAAGT	342	58º	EX5F EX5R	EX5F
Exon 6	F: TGAATGAGTGCCAAGCAAAC R: TTCCCAAAACCCTACAGCAC	277	59º	EX6F EX6R	EX6F
Exon 7	F: GCGAAGGGATGGGTAGGG R: GCATGAGGGGTTTGGTTG	248	58º	EX7F EX7R	EX7F
Exon 8	F: ATCTCCCGAGAGGCTGGGCTGTCT R: CCCGGTCAGGGGATATGAGTCTGT	361	59º	MB30 MB31	EX8F
Exon 9	F: AAGGGGATGGGGAGGCACTCTTG R: CCTCATCTCACCTGCGGGCCA	397	59º	EX9F EX9R	EX9F
Exon 10	F: CCTTGGCCCGCAGGTGAGA R: GTGCTGGGATTACAGGTGCTTTGA	403	62º	EX10F EX10R	EX10F
Exon 11	F: GCCACATTTGGAGTTTGGGGTTC R: AGCAGCTTGGGCTTGTCCCAGA	355	60º	EX11F EX11R	EX11R
Exon 12	F: GGTGCTTTTCTGCTAGGTCC R: TTTTCTGCGTTCATCTTGGCT	347	59º	EX12F EX12R	EX12F
Exon 13	F: CTAGTTGTGGAGAGAGGGTGGC R: GCGGAGTCAGGGCAGGAACGAG	275	60º	EX13F EX13R	EX13F
Exon 14	F: GAAACCTCCTTGTGGAAACTCT R: GAAAAGTATGGTTATCCCGACT	388	58º	EX14F EX14R	EX14F
Exon 15	F: CCAAGGTCATTTGAGACTTTCGT R: GAGAGAAGGTCAGCAAGGGAGTG	388	60º	EX15F EX15R	EX15R
Exon 16	F: GTCCTCTGCCTGCTCCATTTCTT R: ATCCTCCATCTGACCCCTTAGC	350	60º	EX16F EX16R	EX16F
Exon 17	F: GAGCTGGGTCTCTGGTCTCG R: GCGCACAGAAGCATTCACCT	500	60º	EX17F EX17R	EX17F
Exon 18	F: GAGCGGTGGGAAGTGACTGAAT R: TGGTGCCATCTGCTGTTGTGTG	580	59º	EX18F EX18R	EX18F
ApoB (exon 26)	F: GAGCAGTTGACCACAAGCTTAGCTTGGAA R: GGGTGGCTTTGCTTGTATGTTCTCCGT	343	59º	P61 P62	P61
ApoB (exon 29)	F: CCAAGATGAGATCAACACAATC R: AACTTGACTTGAGAGTTGGG	334	59	MB63 MB64	MB63

Table A I.3. *Primers used for the insert of interest automated sequencing.*

Hybridization site	Orientation	5' – 3' Sequence	Primer
LDLR	Forward	CAGCGCTGAGTGCCAGGATGG	MB8
LDLR	Reverse	CTGGCGGGACCACAGGTGAGC	MB9
LDLR	Forward	CCGCAGCGCTGTAGGGGTCTTTAC	MB10
LDLR	Reverse	TGACCAGTCCCGGCAGTCTCTAGC	MB11
LDLR	Forward	ACTGCCGGGACTGGTCAGATGA	MB12
LDLR	Reverse	GCCGGTTGGTGAAGAAGAGGTA	MB13
LDLR	Forward	GCGTGAACCTGGAGGGTGGCTACA	MB14
LDLR	Reverse	GGGGCCTGGATGTCTCTGCTGATG	MB15
LDLR	Forward	GGAACTCCCGCCAAGATCAAGAAA	MB16
LDLR	Reverse	TTGGCTGGGTGAGGTTGTGGAAGA	MB17
LDLR	Forward	TTCAGTGCCAACCGCCTCACAGG	MB18
LDLR	Reverse	TCGGGAACAGGTCGGGTGGTTG	MB19
LDLR	Forward	GGGGCCACCCCTGGGCTCAC	MB20
LDLR	Reverse	AAGGCCGGCGAGGTCTCAGGA	MB21
LDLR	Forward	CACGATGGGAAGTGCATCTCTC	P169
p.cDNA3	Forward	GGGACTTTCCAAAATGTCGTA	pCMV5
p.cDNA3	Reverse	TTTATTAGGAAAGGACAGTGGG	pFGHR1

Appendix II - Production of *LDLR* gene variants

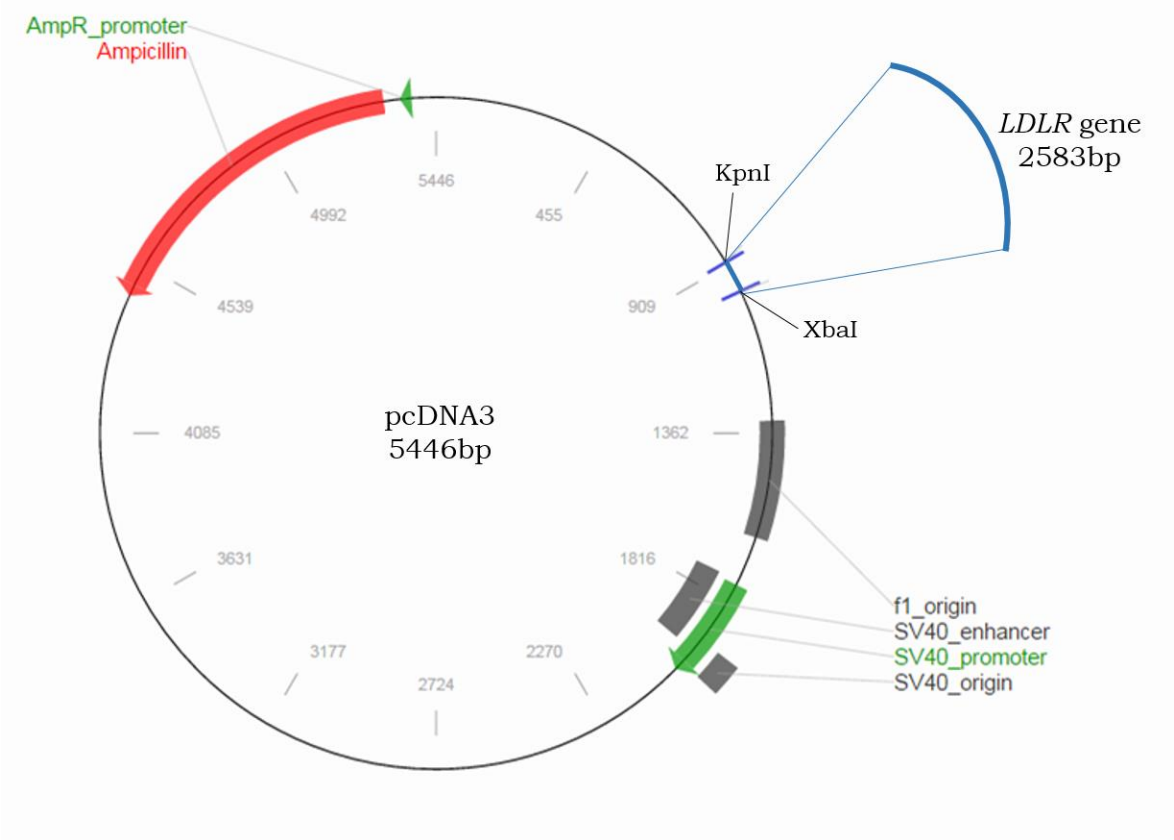


Figure A.II.1. *Simplified representation of pcDNA3 plasmid. Features as the SV40 promoter, the ampicillin resistance gene and the restriction sites for KpnI and XbaI enzymes between which the LDLR cDNA was inserted are represented. Adapted from www.addgene.org/vector-database/2092.*

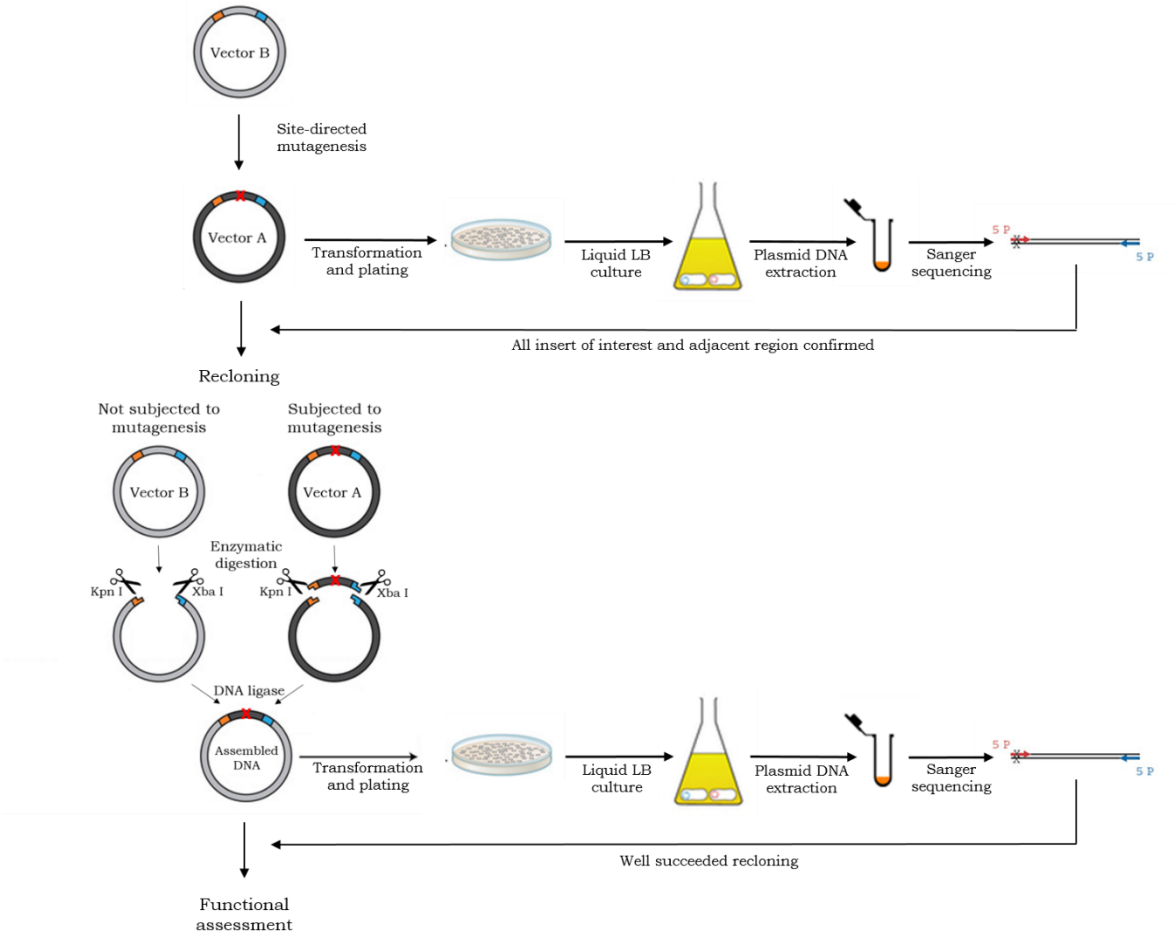


Figure A.II.2. *Simplified scheme of LDLR variants production. After site-directed mutagenesis, all insert of interest and adjacent region was confirmed by automated Sanger sequencing. Both vectors are digested with KpnI and XbaI and the mutated insert of interest is then introduced into the vector B, which was not subjected to mutagenesis, through a ligation reaction. Recloning was also confirmed by automated Sanger sequencing. Adapted from Siu-Hong Chan, Ph.D., New England Biolabs, Inc.; Restriction Endonucleases: Molecular Cloning and Beyond; (https://www.neb.com/products/restriction-endonucleases/restriction-endonucleases/restriction-endonucleases-molecular-cloning-and-beyond).*

Appendix III – Production of *LDLR* gene variants

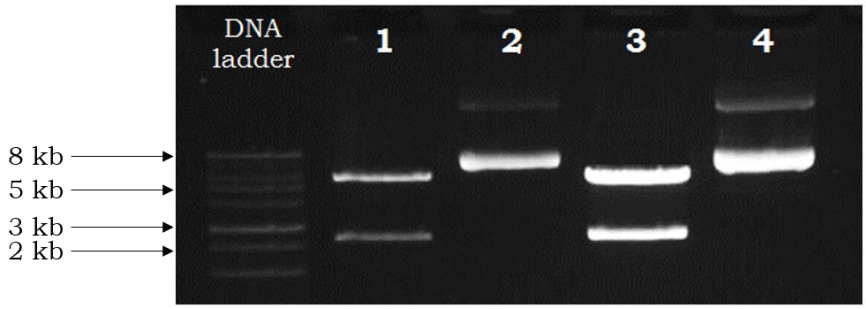


Figure A III.1. *Restriction products after digestion with KpnI and XbaI for 2 h. 1. Digested pcDNA3_LDLR (not subjected to mutagenesis); 2. Undigested pcDNA3_LDLR (not subjected to mutagenesis); 3. Digested pcDNA3_LDLR (subjected to mutagenesis); 4. Undigested pcDNA3_LDLR (subjected to mutagenesis). In lanes 2 and 4 (undigested plasmids), two bands are observable: one corresponding to the relaxed form of plasmid DNA, which migrates slower in the agarose gel due to its conformation, and the supercoiled form with a more intense band. In lanes 1 and 3, two bands were observable – one corresponding to the 5.4 kb of pcDNA3 and the second corresponding to the 2.3 kb of LDLR insert. First band of lane 1 and second band of lane 3 were isolated and purified, in order to be included in the ligation reaction.*

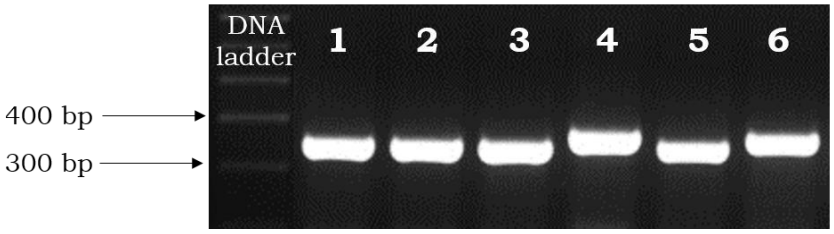


Figure A III.2. *PCR products after amplification with primers MB10 and MB11, which hybridised at a distance of approximately 500 bp in the wt plasmid. When the deletion was present, a difference of 21 bp was observed (lanes 1, 2, 3 and 5), these colonies were used to transform bacteria and the extracted plasmid DNA was sequenced to verify the existence of the desired deletion.*

Appendix IV - Sequence analysis

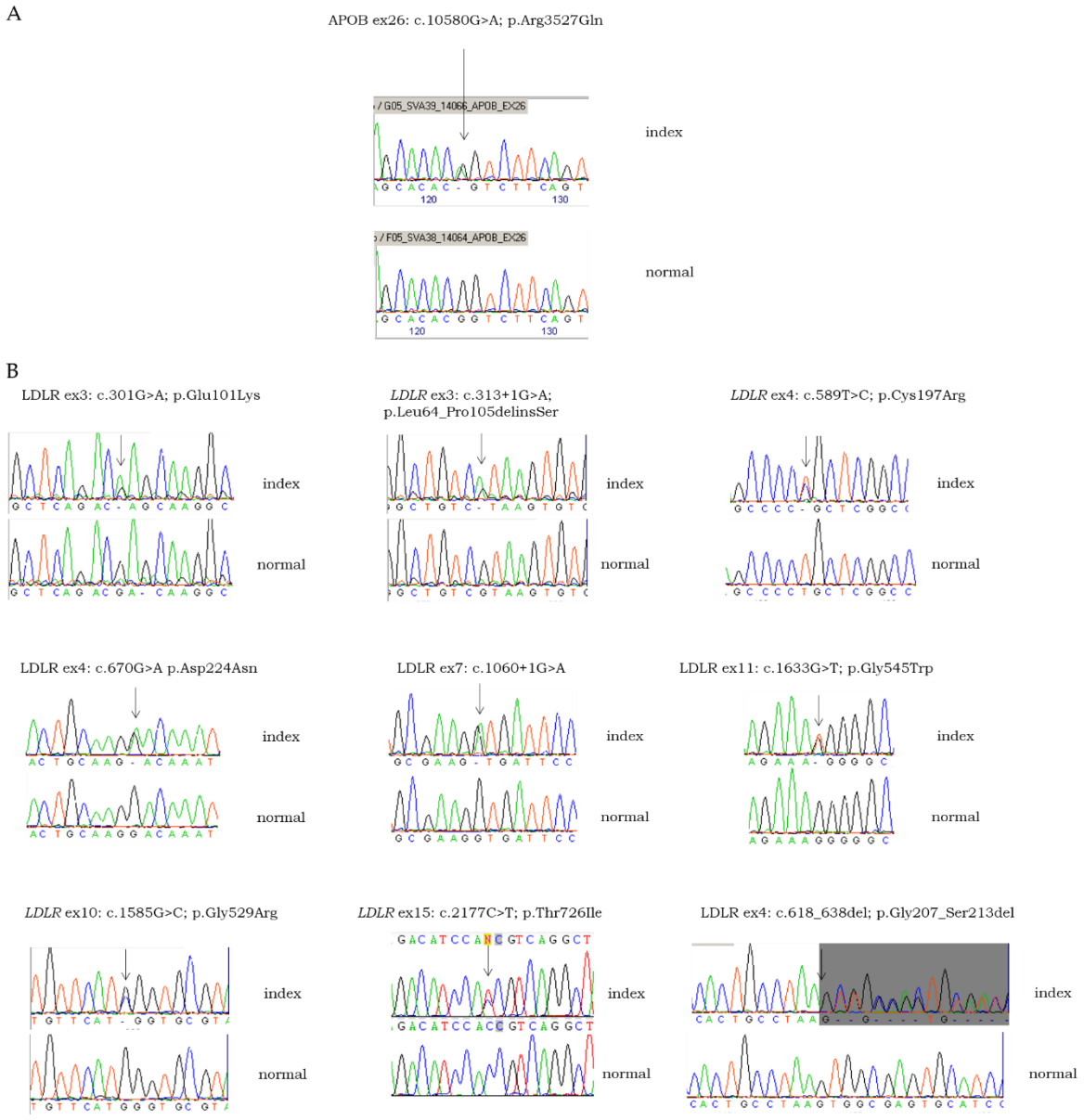


Figure A IV.1. *Electropherograms of variants found during molecular diagnosis. (A) Variant identified in APOB gene; (B) Variants identified in LDLR gene.*